

The smack of firm government

Last week's government paper on British research has the air of finality proper in a policy statement. The discussion, after all, is long since over. The question now is whether the new policy is wise.

THE much advertised (see *Nature*, *ad nauseam*) White Paper on British research policy (see page 385) is intelligently argued. Even so, there are many who will be startled by the tone of the seventh paragraph, which reads:

"...science needs Government and public funds. The decision for Government, when it funds science, as it must, is to judge where to place the balance between the freedom of researchers to follow their own instincts and curiosity, and the guidance of large sums of money towards achieving wider benefits, above all the generation of national prosperity and the improvement of the quality of life."

That leaves no doubt of whom the British government intends to be the boss. To its credit, the document goes on to say that "the Government does not believe that scientists should or could be told, from above, what to work on in order to generate relevant and industrially applicable results". The declared purpose is to create a framework within which "relevant and industrially applicable results" will be the norm.

This is one of several passages which deny the Haldane principle, that governments should fund research but leave the choice of projects to working scientists, but it has been overlooked that Haldane's objectives in 1917 were the same as those now declared. So what has changed? The implicit answer is that "no one nation" can now follow all fields of inquiry. Formally, that is correct. Nevertheless, if the new framework functions as intended, the pattern of research at British universities and research institutes will radically change: "concentration", a slogan since the late 1960s, will bite hard. And to the extent that an unchanged budget (there seems no hope of extra funds) will have to support much relatively expensive applicable research, the pattern of basic research is bound to be transformed. Worse, there will be endless arguments about whether particular projects "underpin" research councils' chosen strategies. Who, for example, will fund the interpretation of quantum mechanics?

Thus everything will depend on how the new system is managed, and by whom. Mr William Waldegrave, the Chancellor of the Duchy of Lancaster (not to be mistaken for the Chancellor of the Exchequer, who lost his job last week), has not fallen for the cruder proposals urged on him last year. He also acknowledges that much of the territory that lies ahead is unexplored. And although the new director-general of the research councils is to be in place by the end of the year, and the research councils operating under

new charters three months later, the document seems to accept that it will take time for new patterns of spending to evolve. If Waldegrave can keep to that resolve, he may carry the research community with him. With the British economy still on the slide, many researchers would be cheered to believe that they are helping. But too much clamour by the new bosses for quick results could be disastrous.

The other joker in the pack is the plan that priorities in research should be in part determined by a technique called technology foresight. The recipe would be more compelling if more people knew what technology foresight means. From the account in the White Paper, it is both a method of forecasting and a means of social engineering, which will be supervised by a steering committee spanning the whole of British science, technology and industry. Panels of experts will collect information about research opportunities in different sectors, make preliminary decisions and put them out for further consultation, in the process seeking to create networks including both creators and exploiters of innovation whose members will develop a sense of kinship in a common enterprise. It is as if the civil service were being asked to make neighbours love one another by organizing sewing circles, or chess clubs, throughout Britain, with the implied understanding that those who show the most affection for each other will be given extra help in their daily lives. It seems a cumbersome and even corruptible substitute for merely 'picking winners'. At this stage, the only certainty is that some business school will have the wit to organize courses on the subject.

In other respects, the White Paper is plainly wrong — or so ominously silent that its intentions are suspect. The issue of pay is such an issue. The White Paper says enough about careers in science to persuade its readers that the government (or at least Waldegrave) cares. The plan that publicly supported PhD courses should be preceded by a one-year MSc course and tough assessment thereupon is sound; it may also help to soften the plight of researchers who live from one short-term contract to another that the new research councils are to try to persuade universities and other institutions to accept some responsibility for them. But it matters more for the health of the research community that these people should be paid decent wages. To say that research councils already deal more generously with students working in industrially linked projects cuts no ice.

The proposed Council on Science and Technology is



also a fudge. Billed as a replacement (before the end of the year) for the Advisory Council on Science and Technology, it will differ only in that Waldegrave himself will be the chairman and that the membership will be even larger than at present. The hope is that it will be "an important forum of largely independent members", but that does not mean that the committee itself will be independent. On the contrary, it is likely to be as susceptible as its predecessor to browbeating by determined ministers and their chief scientific advisers. Waldegrave, who has responsibility for open government as well as for science and technology, is not a browbeater by temperament, but his White Paper will institutionalize his office, and his good manners do not vouch for the behaviour of his successors.

So, flaws apart, should the new document command respect? In reality, there is no choice. As counterpoint to its prescriptive tone, the White Paper runs two other contrasting themes: an upbeat view of the quality of British research and a downbeat reflection of the government's belated realization that the British economy is structurally in a dreadful state. If the second condition continues, research will shrivel anyway. Waldegrave's plan may be the only politically feasible way of keeping a budget of any size. Its most awesome feature is that the changes it will bring will be irreversible unless and until science creates a lot more wealth in Britain. □

Pay no dues yet

Lapsed members should ask for better terms before agreeing to rejoin UNESCO.

MUCH has happened to the UN Educational, Scientific and Cultural Organization (UNESCO) in the eight years since the United States and Britain quit their membership. Most important, Dr Amadou-Mahtar M'Bow has been replaced by Dr Federico Mayor, a distinguished Spanish biochemist. Second, some of M'Bow's madcap schemes, such as that for a "new world information order", have disappeared from the agenda. Moreover, the organization continues many of its good works, support for the International Council of Scientific Unions and for the International Centre for Theoretical Physics at Trieste, for example. So should not the United States and Britain rejoin, as President Bill Clinton has been hinting?

Sadly, and on the terms now on offer, the answer should be, "No, not yet". The reason is that, over the years, UNESCO has neglected tasks it could uniquely perform for the whole of its constituency. It reckons to be something of an expert on copyright, for example, but has been inconspicuous in tackling the question of how (if at all) to extend copyright to electronic texts. It could (and should) have taken a closer interest in the huge movement of intellectuals (from one country to another) now under way, predominantly from Central Europe and points further East. And while it is proper that much of its attention should be directed towards developing countries, the lasting benefits

of its multinational research collaborations could often be won, at a fraction of the cost, by less ambitious bilateral schemes. □

Conference unwanted

The annual AIDS conferences should be replaced by more effective channels of communication.

THE Ninth International AIDS Conference in Berlin this week appears to be following the pattern established by its predecessors. A substantial proportion of the world's full-time AIDS researchers will be there, as will be health-care administrators, representatives of AIDS patients (some of whom will attend in person) and a small army of journalists. For those who wish to catch public attention, say for an untested candidate vaccine, the announcement of a research collaboration or the denunciation of the "AIDS hypothesis", Berlin will be splendid venue. None of this implies that no serious science will be done, but merely that Dr John Moore is right to argue (on page 391) that the AIDS conferences have outlived their usefulness, and that they should be stopped.

The origin of this conference series should not be forgotten. Those who planned the first event in the mid-1980s did so at a time when the tragic nature of the infection was heavy on their minds, but when it also seemed that there would be rapid progress towards the identification of the virus and the prophylaxis of the disease. So what more natural than that the research community should resolve to report back at regular intervals in a setting in which the AIDS community would also pay a part? The expectation, no doubt, was that there would be a cheerful tale to tell in a not-too-distant future.

Seven years have gone by since the first conference, and the implied promise has already been broken. The biomedical research community cannot in logic be blamed that the AIDS virus has turned out to be a more obdurate agent than originally expected, but it is blamed. It would be in the public interest if the disappointment were more openly acknowledged. Then the research community would be less easily pilloried than at present, which would make for more seemly open conferences. It would also then be possible for bystanders more easily to pick out the directions that offer hope of better treatment. But global conferences like this week's would also then be less informative than the links the AIDS community and its researchers have established with physicians and researchers.

For the time being, anyway, the chief public (as distinct from professional) interest has shifted from biomedical affairs to the interaction between behaviour and infection. Professor Roy Anderson's article on page 393 should give the lie to those who claim that heterosexual AIDS has magically ceased to be a problem. In passing, it confirms that changes in behaviour have already been influential. Does one have to travel to Berlin to hear that? □

Biomedical trust fund is proposed to support basic research at NIH

Washington. Two prominent US senators have proposed a trust fund that would provide billions of dollars a year for biomedical research from a tax on each insurance policy written under any reform of the US health care system.

More than a hundred professional societies and disease support groups have endorsed the idea, excited by the thought of more money for their members and pleased that it places research at the centre of attempts to reduce the cost of health care, which this year will exceed \$900 billion. However, the proposal appears at this point to be more of an angry response to the budget that President Bill Clinton has proposed for the US National Institutes of Health (NIH) than a comprehensive, long-term approach to improving public health.

At a press conference last week to announce their proposal, Senators Tom Harkin (Democrat, Iowa) and Mark Hatfield (Republican, Oregon) described the failure of biomedical research funding to keep pace with the health needs of the US public. The trust fund, they said, could raise NIH's current annual budget of \$10.3 billion by as much as 50 per cent.

"Until we adopt this approach", says Harkin, who is chairman of the Senate appropriations subcommittee that reviews the NIH budget, "the health care debate will be about paying bills, not preventing them". Adds Hatfield, "the ultimate cost savings in health-care reform is prevention, and the best approach is through research".

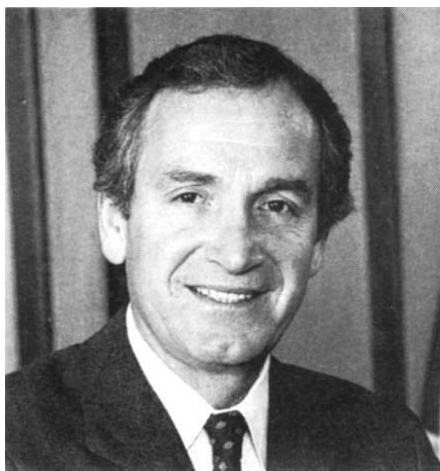
The trust fund would be derived from a \$5 tax on each health insurance premium paid under any new health care system and would supplement the annual NIH appropriations, explained Hatfield, the senior minority member of the appropriations subcommittee. That goal could be met with a provision that the trust fund would be tapped only after each institute received a certain minimum level of funding, adjusted annually to account for inflation.

Testifying the previous day before Harkin's subcommittee, outgoing NIH director Bernadine Healy said that she would like to see the money used to raise the number of grants awarded each year from a current level of 21,500 to about 32,000. The infusion of money would also increase the chance of having a worthwhile idea funded, she said, raising success rates from a projected 21 per cent in 1994 to 33 per cent by 1997.

"This would be a signal to the community that we had reached a sustainable level of support that can meet the tremendous opportunities we face", she said. Any addi-

tional money, she added, "would focus on neglected areas such as career training and the growing problem of aging research facilities".

Healy also disclosed that her initial 1994



Harkin puts his trust in research.

budget request was \$12.3 billion, a \$2 billion increase endorsed by her political bosses at the Department of Health and Human Services but ignored by the new Clinton administration. "We proposed two years of 18 per cent increases, followed by increases

of 14 per cent and 6 per cent", she said. "And for the first time in its history, the department accepted our full request and passed it along to OMB [the White House Office of Management and Budget]."

Several months later, when the new administration examined NIH's request, she said, the agency was told to develop a budget that matched a projected 4 per cent rate of inflation, which was later amended to a 3.2 per cent, across-the-board increase that made exceptions for a handful of programmes. "We were not even at the table when the final budget was put together", she added, "and we weren't allowed to appeal. Instead, we were told how much to put into each special initiative — for AIDS, breast cancer, high-performance computing and advanced materials."

Harkin and Hatfield say that they are not wedded to a tax of a specific size or even to the idea of a trust fund itself. Instead, they say that their goal is to make sure that the task force assembling the reform plan, led by the president's wife, Hillary Rodham Clinton, recognizes the importance of biomedical research and gives it a prominent place in its final proposal. Harkin said that Clinton and his aides had no specific reaction when he described the idea to them in a series of discussions over the past two months.

Jeffrey Mervis

Life sciences take a dive

Washington. The life sciences have been demoted within the US Office of Science and Technology Policy (OSTP), a move that has upset biomedical researchers and even caught the eye of an influential Democratic senator.

John Gibbons, director of OSTP and presidential science adviser, has reorganized his office to give greater attention to environmental issues, and Robert Watson, a NASA scientist, is preparing to join the office as the new associate director for environment. But because OSTP is allowed only four associate directors, one of the existing posts had to be refigured, and the life sciences lost out. That outcome is not new to those in the discipline: historically, the science adviser is a physicist and the office concentrates on matters of national security and technology.

Under the new alignment, the associate directors for the physical sciences and the life sciences will be combined under a single associate director for science. Gibbons says that he is looking for someone with broad

research experience to fill the science post, and that the life sciences and the physical sciences will be handled by assistant directors within that office.

A third associate directorship, for technology, has been filled by Skip Johns, whom Gibbons brought with him from the congressional Office of Technology Assessment. The fourth associate director, not yet named, will cover national security and international affairs.

The restructuring is sure to displease Senator Tom Harkin (Democrat, Iowa), the chairman of the appropriations subcommittee that reviews the NIH budget and an advocate for greater spending on biomedical research (see above). Last week, at a hearing on the 1994 budget for the National Institutes of Health (NIH), Harkin talked about the need "to elevate our voices" and asked if there was a life scientist "at the elbow" of the science adviser, who is a physicist. "It is my understanding that there is not", replied Bernadine Healy, director of NIH.

Jeffrey Mervis

ESO negotiates with Chile over agreement on telescopes

Santiago & Munich. The European Southern Observatory (ESO) is renegotiating a 29-year-old treaty with Chile, host country for its 15 telescopes at La Silla, in the hope of reaching an agreement that will not disrupt the construction of the Very Large Telescope (VLT) at a second site in the northern Chilean desert. This month the two sides will meet in Santiago

review process and reduce the scientific return from the instruments. Officials say that the 25–30 active Chilean astronomers are too few to make good use of the time — which would amount to 18 days every six months — and that its member countries, each with many more potential users, already receive less time than they would like.

The negotiations are the result of an investigation by the lower house of parliament into the acquisition of land for the VLT. Last month, legislators voted unanimously to support its conclusion that the ruling junta donated the land illegally and declared that parliament must give its approval to any future transaction. The legislators also endorsed the idea of a

10 per cent share of the viewing time and threatened to terminate the treaty unless the new terms are more favourable. The government has not as yet responded to the report.

ESO says that it acquired its 700 square km atop Paranal legally and that any blame rests with the previous government headed by General Augusto Pinochet. In fact, says ESO's spokesman Richard West, "we bought the land twice", because it was easier to settle a belated claim from a second person rather than taking the matter to court. "This shows a little about the confusion of who owns what in Chile", says West.

The first unit of the VLT should be ready in 1996, with the entire package of four, 8-meter mirrors completed around 2001.

In a related matter, ESO is wrestling with allegations of unfair labour practices brought by a trade union formed in 1991 for Chileans working at La Silla. ESO, which offers its European employees higher pay and better benefits as an inducement to work in Chile, recognizes only the existing union formed 25 years ago.

Meeting this week, the ESO council was expected to prepare a negotiating stance for the next session on 18 June. Chilean scientists are optimistic that ESO's new director, Riccardo Giacconi, will prove to be more flexible than his predecessors and that an agreement can be reached before the end of the year. A new president of Chile will be elected in December, and the legislature is expected to adjourn several weeks beforehand.

Jeffrey Mervis & Alison Abbott

US asked to help in reunification of German science

Washington. Germany wants US scientists to take part in collaborations to help rebuild the research base of the former East Germany. "We hope that the United States will show as much goodwill to East Germany now as it did to West Germany after 1945", said Gebhard Ziller, the top civil servant in the German Federal Ministry of Research and Technology, addressing some 300 scientists and industrialists meeting last week at the US National Academy of Sciences.

Ziller concedes that deals between American researchers and their counterparts in East Germany will take time and that the first priority is to make contacts. But Robert Nowak, chief scientist for Dow Chemical of Midland, Michigan, expressed concern that not enough large US companies have shown any interest in such research collaborations. "I think we're missing the boat", he said.

German officials hope that the new German-American Academic Council (see *Nature* 362, 779; 1993) will provide a "major stimulus" for such collaboration after its 30 members are selected and meet later this year. But no US money has yet been made available to help fund the council, which has been promised DM18 million (US\$11.5 million) over the next four years from the German research ministry.

The restructuring of research institutes and universities in the new *Länder* has resulted in a hundred new establishments and research teams, each keen to pursue international collaboration. But officials concede that the situation in industrial research is "alarming": a work force still 90 per cent state-subsidized has already been slashed from 100,000 to 15,000 and is expected to fall further.

Kurt Biedenkopf, prime minister of Saxony and one of the most respected politicians from eastern Germany, painted a grim picture of academic institutions and said that scientists and engineers had been treated more favourably by the old regime than other academic sectors. "The really important thing is the rebuilding of the humanities, because there is nothing", he said. "The real devastation took place in the heads of people."

Colin Macilwain

Historic vote

London. The first contested presidential election in the 205-year history of the Linnean Society of London has been won by palaeontologist Brian Gardiner of King's College, London. The challenger, geneticist Bryan Clark of the University of Nottingham, received 33 votes to Gardiner's 75 in last week's election.

H.G.

"We want to be shown the respect that a host country deserves."

Maria-Tereza Ruiz
University of Chile astronomer

for the second time as they work against a deadline of 1 September set by the Chilean parliament when it asked for the 1964 treaty to be renegotiated.

Chile wants to heal a festering wound to its national pride inflicted during an era when its politicians were flattered by any attention from Europe. That attitude has changed since the military dictatorship was replaced in 1990 by a civilian government and Chile began to reassess its links with the rest of the world.

Most Chilean astronomers believe that the current agreement fails to serve them well, and that it must be modified to give Chile a share of viewing time on the largest ESO telescopes and comparable pay and career opportunities for technical and scientific staff from Chile. They would also like ESO to support programmes to increase the size of the astronomy community and to become more active in public education.

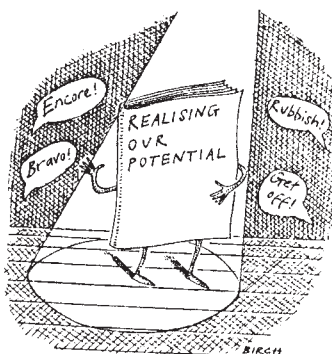
"We have taken the position that it is reasonable for Chile to receive a minimum amount of time on the ESO telescopes and that Chilean science should benefit from its existence", says biochemist Jorge Allende, president of the Chilean Academy of Sciences. Adds Maria-Tereza Ruiz, an astronomer at the University of Chile, "we want an agreement that satisfies both sides. We want to have the VLT in Chile, and we also want to be shown the respect that a host country deserves."

ESO has historically resisted the idea of any guarantee of viewing time (the most commonly cited figure is 10 per cent) on the grounds that it would subvert the peer

Britain cheers and jeers at a *status quo* White Paper

London. The establishment is happy; its critics are not. That, in a nutshell, summarizes reaction to the British government's White Paper (policy document) on the organization of science and technology, *Realising our Potential*, published last week in London.

The lack of radical change has reassured those who feared that William Waldegrave, the Chancellor of the Duchy of Lancaster and as such the cabinet minister responsible for science, would use the report to re-



distribute the government's responsibilities for research.

Universities, for example, are pleased with the decision to maintain the dual

support system. The research councils are similarly relieved that Waldegrave rejected a suggestion by the Advisory Council on Science and Technology (ACOST) to privatize many of their research institutes and separate the funding of basic and applied research (see *Nature* 360, 614; 1992).

Ironically, the Medical Research Council, which would have been most affected by this plan and was one of its sharpest critics,

emerges as a model for the other research councils since it already has a part-time chairman from industry and a full-time chief executive officer. "I am very pleased with the White Paper", says MRC secretary Dai Rees. "The implementation of ACOST's proposals would have been a loss to the nation."

Tom Blundell, the chairman of the Agricultural and Food Research Council, which will be expanded into a new Biotechnology and Biology Research Council (BBRC), says that he welcomes the shift in emphasis as the 'coming of age' of biology-based industry. Industry itself appears pleased with its new 'partnership' role, heavily influenced by input from Michael Heseltine, the minister responsible for the Department of Trade and Industry (DTI).

This role will be achieved, for example through the presence of prominent industrialists on the new Council of Science and Technology, which will replace ACOST, and their increased role in recommending priorities for the research councils. DTI officials describe the White Paper as a 'victory' for Heseltine's approach, while Fiona Steele of the Confederation of British Industry says simply, "we like it".

In contrast, those who had been pushing Waldegrave for more radical changes are frustrated over, for example, the lack of any commitment to significant increases in funding for science, or of any promise to shift spending on military research and development into the civilian sector.

"The absence of either of these is a great disappointment", says John Mulvey of the department of physics at the University of Oxford, secretary of the Save British Science campaign. Mulvey says that pressure on research councils to emphasize their role in wealth creation and to focus on priorities decided by industry "is a considerable lurch to more state control".

Trade unions representing scientific staff have criticized the White Paper's declared intention to continue the privatization of government-funded research laboratories. "We welcome the White Paper's proposals to be more open about setting priorities for science, but we don't like the dogma about privatization," says Valerie Ellis, assistant general sec-

A global search

Britain's Department of Trade and Industry is to establish an international network of technology counsellors to gather information about new scientific and technical advances that could be of benefit to British industry. The initiative is one of a number of developments announced last week by Michael Heseltine, the minister for the department, to coincide with the publication of the government's White Paper.

"Our role is to help British industry make better use of research," says Edward Leigh, who was until last week a junior minister for trade and industry, and actively engaged in challenging efforts by the Foreign Office to cut back support for science attachés working in British embassies abroad (see *Nature* 363, 101; 1993). "Since only a small proportion of the world's R&D is carried out in Britain, it makes sense to find out more about what the rest of the world is doing." **D.D.**

retary of the Institution of Professional Managers and Specialists. The union's members include 90,000 scientists and other specialists, many of whose jobs could be threatened.

The opposition Labour Party has been even stronger in its condemnation. Science spokesman Lewis Moonie described the White Paper as a "complete failure" that condemns Britain to "hobbling into the future with an underfunded science base".

Reaction is mixed to the White Paper's proposals for improving the career prospects for scientists. Many have applauded the introduction of a one-year MSc for graduates before embarking on a PhD — "a welcome move towards US practice", says Sir Martin Rees, professor of astronomy at the University of Cambridge — as well as new measures to encourage universities to provide scientists with more stable careers.

But there has also been criticism of the government's intention to introduce a more careful selection of PhD candidates, leading to fewer graduates studying for such degrees. While the research councils see this as a better use of scarce resources, David Triesman, president of the Association of University Teachers, says that it "will give the wrong signal to our brightest students".

There is also concern about the impact on research of splitting the Science and Engineering Research Council (SERC) into two bodies: an Engineering and Physical Science Research Council and a Particle Physics and Astronomy Research Council (PPARC). In particular, some fear that the increased exposure of "big science" to political influence could make it more vulnerable to attempts at seeking quick reductions in the science budget. **David Dickson**

Waldegrave gets help

The British government has replaced a classicist with a science graduate and former business manager to serve as junior minister to William Waldegrave, the cabinet minister responsible for science. Part of his job will be to help implement plans for a reorganization.

In last week's cabinet reshuffle, David Davis, who studied molecular sciences and computer science at the University of Warwick and worked for the sugar company Tate and Lyle before entering Parliament in 1987, replaced Oxford-educated Robert Jackson as Parliamentary Under-secretary at the Office of Public Service and Science. Another casualty of the reshuffle was Edward Leigh, a junior minister in the Department of Trade and Industry (DTI).

Attention in Whitehall is now turning to the crucial post of director-general of research councils. Two potential candidates are Sir Mark Richmond, who will lose his position as chairman of the Science and Engineering Research Council when the council is split in three later this year, and Geoffrey Robinson, chief scientist at the DTI.

D.D.

Soccer boom may score for Japanese science

Tokyo. Will Japan's sudden enthusiasm for soccer, marked by the start last month of Japan's first professional soccer league, help to break through the rigid ceiling on government spending for science in the universities? The basis for that seemingly odd question is a history of using part of the proceeds

are run by the Japan Racing Association, to support science. For example, the ministry's new rice genome project gets a large proportion of its funding from this source (see *Nature* 356, 181; 1992). The Ministry of Transport also feeds some of the profits from betting on motorboat races under its jurisdiction into the development of technology.

Keirin provides about ¥35 billion (US\$315 million) a year or the equivalent of about 12 per cent of MITI's overall research and development budget. These extra sources of funds have given these ministries a way to fund projects that otherwise might have been impossible because of the squeeze on government expenditures from Japan's huge national government debt. But the Ministry

of Education, Science and Culture (MESC), has no such source of funds, and the universities and the university research system under MESC have suffered greatly during

the last decade of fiscal restraint. This is where soccer might help.

It is widely anticipated that the proposed football pools system will be overseen by MESC; the ministry is already responsible for high school soccer teams. Furthermore, Kosuke Hori, a former minister of education, science and culture, who has been seeking ways to break through the ceiling on MESC's budget, is leading an LDP team looking into ways of promoting sports, including the proposed soccer pools.

But the public is increasingly opposed to the idea, with parents worrying that children might be led astray by betting on football and the leading opposition party, the Social Democratic Party, officially opposed to the idea. To counter that sentiment, one politician from the New Sports Party returned from a visit to Italy expressing the view that filling in football pool forms requires thought and intelligence and could be educational for children.

Leading academics are also sceptical that universities are prepared to accept funds, no matter how indirect, from such a source. "It's easier for national institutes to accept such money", says the former president of one leading national university. But given the reluctance of the powerful Ministry of Finance to increase MESC's budget, soccer may offer an irresistible scoring opportunity for science.

David Swinbanks



Japan gets a kick out of its first professional soccer league.

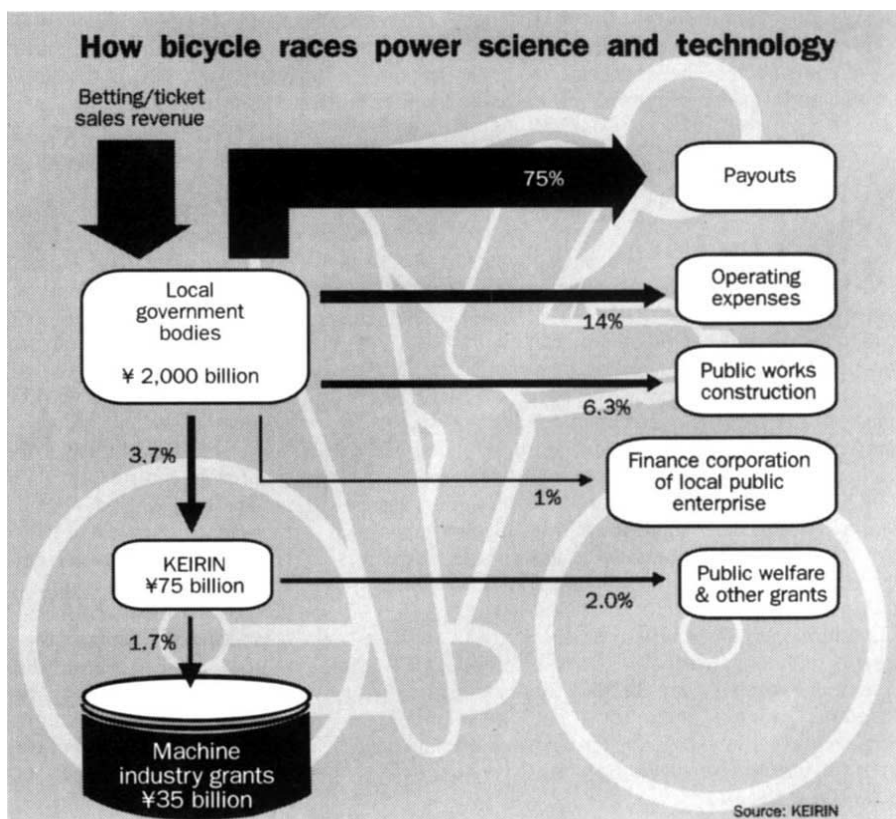
from gambling on sports to support science and technology.

The popularity of soccer has skyrocketed after the arrival of famous Western soccer stars, like Britain's Gary Lineker, to play on teams sponsored by companies and local governments. The creation of the new J-league has prompted members of Japan's ruling Liberal Democratic Party (LDP) to consider a football pool betting system, like that in Britain and Italy, to finance improvements to the social infrastructure.

At present, the LDP politicians promoting the scheme are suggesting only that profits be fed into the support of sports, in particular soccer. But past experience suggests that some of the money may be diverted into science and technology.

In 1957, the postwar popularity of bicycle races led local governments and the Ministry of International Trade and Industry (MITI) to establish the Japan Keirin Association, an organization that funnelled some of the huge profits from betting on bicycle races into the bicycle industry as well as into the sport itself. Over the years, the range of industries and activities supported by Keirin have expanded to the point where the money frequently supports government-industry research projects and international conferences in Japan in frontier areas of science and technology (see right).

Similarly, the Ministry of Agriculture, Forestry and Fisheries has used some of the profits from gambling on horse races, which



Races grease the wheels of science

Tokyo. Scientists coming to Japan for conferences on neural networks and biomaterials are sometimes surprised to discover small print in their conference programme explaining that the event is supported in part from the profits of bicycle races.

Conferences are just one of many activities related to the development of future technology undertaken by Keirin, an association overseen by the Ministry of International Trade and Industry (MITI) that distributes some of the huge profits arising from betting on bicycle races.

Bicycle races are held almost every day in Japan at 50 locations, and bets can be placed at the race tracks or by telephone by registered users anywhere in the country. About ¥2,000 billion (US\$18 billion) each year is spent on ticket sales and bets, and the total continues to rise.

When Keirin was established in 1948 its role was to support the bicycle industry, but in the late 1950s it began to support a variety of machine industries. Now a significant part of Keirin's annual budget of about ¥35 billion (US\$315 million) for machine industry grants supports the development of future technologies.

For example, Keirin helps to support the Research and Development Organization for Future Electron Devices, an association of 25 companies set up by MITI in 1981. It also gives about ¥2 billion a year to the Japan Industrial Policy Research Institute, a non-profit organization that disperses small grants and contracts.

D.S.

German state challenges faculty reappointments

Munich. A German court has given Armin Ermisch his job back at Leipzig University, but the cell biologist sits at home because state government officials refuse to accept the court directives.

Hans Joachim Meyer, research minister for the east German *Länder* of Saxony, is challenging court orders to reinstate 13 professors sacked last year for allegedly collaborating with Stasi, the secret police for the former DDR. The order will be challenged as a matter of principle, Meyer says, although no new evidence exists.



Armin Ermisch

(see *Nature* 359, 762; 1992). In Saxony, more than 5,000 academic staff were sacked from the chronically overstuffed universities. Most were simply declared surplus to the new requirements, but 989 were sacked on the recommendation of the Personal-kommissionen, groups of locally selected people set up to assess the 'personal integrity' of academics.

The democracy of the decisions, final responsibility for which rests with Saxony's

research minister, was widely challenged. Many of the victims felt that the Personal-kommissionen were influenced by personal motives, and 303 took their cases to court.

A tally on the cases, most of which were heard earlier this year, is difficult because no central records are maintained. However, it is believed that most reached a settlement, without any conclusion about whether the allegations were true.

However in 13 cases the courts clearly ruled in favour of the plaintiffs, finding no evidence to suggest that the accused had in any way abused their posts in the Communist Party to the detriment of their colleagues or their students. The government was ordered to reinstate these professors.

Meyer says he will continue to accept the opinions of the Personal-kommissionen against those of the court as long as he has the legal right of challenge. He denies that he is being inflexible, saying instead that he changed his mind about many dismissals when more evidence came to light (the Personal-kommissionen had originally recommended dismissal of 1,651 staff on the grounds of 'personal integrity').

"Since 1968, academic institutions have become instruments of party rule, and this has to change", he says. "I'm not acting as a judge or handing out sentences. I'm simply acting as the cabinet minister responsible for [these changes]."

Ermisch lost his university post at the end of last year but was put back on the payroll in February after a favourable ruling by the labour court in Dresden. But the university, on Meyer's recommendation, asked biology director Karl Drössler to reach "a sensible compromise" with Ermisch (as with all the reinstated professors) in the run up to Meyer's appeals to the *Länd* labour court. "We agreed that professor Ermisch should work mostly at home while the trial was still going on", says Drössler. But Ermisch finds this solution unsatisfactory because he is able to work only on material brought to him by his students. Wolf-Dietrich Arnold of Leipzig University's orthopaedic clinic is in the same position.

Friedel Schnur, a lawyer representing three of the 13 professors, says that Meyer is just being stubborn. "The court ruled in our favour and Meyer has nothing new to say", he says. "We don't expect him to win his appeal, but if he does, we will counterappeal."

If the *Länd* labour court decides that the Dresden court had the right to overturn the decision of Meyer and his Personal-kommission, the minister can appeal to the federal court. The *Länd* court will decide whether or not to allow the appeal.

Alison Abbott

SmithKline backs sequencing company

Washington. Human Genome Sciences Inc. (HGS) of Rockville, Maryland — formed last year to commercialize products from the large-scale gene sequencing programme at the nearby Institute for Genomic Research (see *Nature* 358, 95; 1992) — has formed a strategic alliance with the Anglo-American pharmaceutical company SmithKline Beecham (SKB).

The agreement, believed to be worth upwards of \$100 million, gives SKB the exclusive worldwide rights to diagnostic, therapeutic and vaccine products and services developed from the gene sequencing data generated through the collaboration. In return, HGS will receive royalty payments on the sales of any product developed in the human and animal health sectors as well as an equity investment by the drug company. HGS will, however, retain the rights to use

its gene sequencing data in gene therapy and for the development of therapeutic agents based on antisense technology.

Although the collaboration is unlikely to produce products in the near-term, Stewart Adkins, an analyst with Shearson Lehman, believes SKB sees it as a "long-term opportunity... worth funding because there are so many things that could come out of it".

One week before the agreement was reached, HGS announced the hiring of William A. Haseltine as chief executive officer and chairman of the board. Haseltine, who has taken a two-year leave of absence from the Dana-Farber Cancer Institute in Boston, says that HGS had been in discussions with a number of other pharmaceutical companies and that he is "delighted that we [HGS] have formed an alliance" with SKB.

Diane Gershon

AIDS activists says basic research is underfunded

Washington. Despite extra resources from the US government to combat AIDS, research is skewed against critical basic work on the pathogenesis of the disease and leading researchers are "chronically demoralized", according to a new report by an influential AIDS activist group.

The Treatment Action Group (TAG) — emboldened by its success last year in framing recommendations which Congress last month incorporated into reform of the National Institutes of Health (NIH)'s Office of AIDS Research — plans a year-long campaign for greater support of such work as well as other aspects of its report, to be published next week at the Ninth International Conference on AIDS in Berlin. The report, based on interviews with 36 AIDS investigators in the United States, also calls for a further shift in research away from test-tube work towards investigations that deal with patients.

TAG would like NIH to move substantially away from *in vitro* work and towards *in vivo* research exploring the behaviour of HIV (human immunodeficiency virus) in the human body. "AIDS research came of age at the same time as molecular biology, in the late 1980s" says the report's author, Gregg Gonsalves of TAG. "As a result, people moved away from looking at the patient."

Recent work such as that by NIH's Anthony Fauci on the lymph node (*Nature* 362, 355; 1993) reflects a welcome shift back

towards the study of the pathogenesis of the disease in the body, TAG says. The work should be extended to ways in which the human immune system controls the virus, to find out, for example, why some people with high CD4 counts contract AIDS while others register low counts but show no symptoms, and why those directly exposed to HIV manage to resist infection.

Gonsalves estimates that NIH receives two to three times more top-quality grant applications for basic work on AIDS than it can fund. Although he concedes that a similar situation exists in most fields, he would like work on AIDS pathogenesis to be made an exception because "it is our number one public health problem".

TAG argues that the best way to find a cure for AIDS is to put more of the \$1.3 billion that the US government spends on AIDS into investigator-initiated individual grants known as R01 awards. "We don't need a Manhattan project for AIDS", says Gonsalves. "The best thing to do is to increase the size of the R01 pool." He argues that the ability of large research centres to marshal support for large clinical trials, for example, has given them an unfair advantage over individual researchers desperate for funds.

"We were disturbed to see the level of morale in the research community", Gonsalves says. "Funding is very tight, and there is chronic demoralization among the researchers."

Colin Macilwain

Australian state wants claim to native plants

Sydney. The government of the Australian state of Queensland is expected some time this year to declare its right to a share of any financial gain made from research involving native plants and animals.

The legislation, which has yet to be written or submitted to the Queensland Parliament, will state that the government is entitled to part of the money from the sale of compounds or genes derived from native plants and animals. It is believed to be the first attempt by any government to assert a blanket legal claim to the commercial rights resulting from such research.

Although the policy, to be attached to the existing Nature Conservation Act, may well have little effect outside Australia, or even outside the state, its purpose is to reinforce the state's international rights to naturally occurring compounds and genes under the Biodiversity Convention. That convention was signed in June 1992 and is intended to balance the rights of developing nations with the commercial practices of the industrial

world. Although the federal government is responsible for international matters, states are free to make laws on matters such as logging in forests and mineral exploitation.

Gordon Guymer, head of the Queensland Herbarium (a Queensland government department) and a proponent of state claims to its natural resources, points out that a compound found in the seeds of the native tree, the Morton Bay Chestnut, which has shown promise as an anti-cancer agent, has recently been patented by a Japanese company. A better-known example is the discovery that the bark of the Pacific yew tree, native to Queensland as well as to other parts of the world, contains taxol, a substance effective against ovarian cancer.

An official in the office of the chief scientist for the federal government says that he is puzzled by the proposed legislation. Although it is possible to patent genetically altered animals, he says, it is another matter to declare one's rights to natural life forms.

Mark Lawson

Rumoured good news raises share price of vaccine company

Washington. Shares in the US biotechnology company Immune Response Corporation have doubled in value over the past few weeks in anticipation of the results of therapeutic AIDS vaccine trials to be announced next week at the international AIDS conference in Berlin.

But as interest in the stock has soared — in one day last week its price gained \$4, to \$22, and one-sixteenth of its shares changed hands — some AIDS researchers were critical of the publicity surrounding the release of the closely guarded results and accused the company, which is based in San Diego, California, of trying to put pressure on the US Food and Drug Administration (FDA) to approve its vaccine.

The vaccine used in the trials was developed by Jonas Salk, a co-founder and director of the company and chairman of its scientific advisory board. Salk, who holds 1 per cent of the company's 16.5 million shares, is a champion of 'killed-virus' vaccines of the type that he invented to treat polio before a successful alternative was devised by Albert Sabin.

But some AIDS researchers who doubt the value of therapeutic vaccines in general, and a killed-virus vaccine in particular, are worried that the significance of the trial results will be exaggerated by thousands of reporters in Berlin looking for miracle-cure stories. Salk's vaccine excludes gp120, the basis of other therapeutic vaccines under development. "I'm sceptical about therapeutic vaccines — the jury is still out on their potential", says Paul Maddon, chief executive officer of the New York biotechnology company Progenics Pharmaceuticals Inc., which is developing diagnostics, therapeutic treatments and preventative vaccines for HIV.

Salk declined to comment on the significance of the results of the trials, which will be announced on 9 June by the company's chief scientist, Dennis Carlo. The results involve a five-year follow-up to a study of 23 patients, as well as preliminary data from a one-year trial of 103 patients.

The FDA has been reluctant to license killed-virus vaccines because of the possibility of the accidental infection of patients if the vaccine is mishandled. But market analysts speculate that the company could file an application for the vaccine within months.

Anthony Fauci, director of the National Institutes of Allergy and Infectious Diseases and of NIH's Office of AIDS Research, declined to comment specifically on the Salk vaccine. But he pointed out that "the history of the international AIDS conference is that things do get blown out of proportion".

Colin Macilwain

Science and religion

SIR — Brian Josephson's defence (*Nature* 362, 583; 1993) of a scientific analysis of religion is a well intentioned attempt at a non-partisan approach to understanding the phenomenon of 'faith'. But are the resources and facilities of even the Cavendish Laboratory sufficient for the task?

Josephson unfortunately reveals his own professional presumptions. He claims that "any scientific study of religion should . . . take account of the fact that a central theme of religion . . . is the attempt to maximize 'human goodness'", and then bases his research proposal on that assumption. Although I personally doubt that gene therapists will ever be able to isolate and 'turn on' a (sadly too) latent 'human goodness gene', such thinking illustrates the strange hybrid of behaviourism and Platonism that characterizes much discussion of the 'scientific' analyses of 'religion'.

The central theme of most of the major world religions is not maximizing goodness *per se*, but achieving some given state of salvation (whether that is the Christian or Islamic notion of resurrection, or the Buddhist notion of nirvana). Religious systems and practices are understood by the faithful as simply human structures developed to teach subsequent generations about what they take to be an actual (future) physical reality. At most, maximizing human goodness is only a means to this end.

In the wake of the siege and fire at the headquarters of a religious sect in Waco, Texas, in April, when many people died, it would be too easy for the cynic to dismiss people with such strongly held metaphysical beliefs in salvation simply as suffering from dangerous collective delusions. No doubt Richard Dawkins finds some empirical support for his 'religious virus' theory in the events at Waco. However, it must be remembered that people do not commit their lives merely to received and comforting hypotheses of an afterlife (or a non-afterlife in the case of Buddhism). An individual's commitment is also based (phenomenologically) on his or her own complex personal experiences of life. Newton, Faraday and a great many contemporary scientists (and presumably many at the Waco compound too), would claim to have directly experienced a purposive, nurturing and transpersonal providence, which they have chosen to call the will of God. Dawkins' deeply held atheism, similarly, is due only to his own non-experience, or non-recognition, of such a phenomenon, resulting in consequent inability to frame an acceptable God hypothesis. Certainly psychopathological factors may affect a particular individual's response to such a phenomenon, or confuse his or her under-

standing of the relationship of this phenomenon to God. Hence there is a very real need for doctrinal (non-cultic) approaches to these notions.

The scientific interpretation of physical phenomena has a vital contribution to make in the systematic analysis of the phenomenological experiences entailed in religious belief. Unfortunately, science is still considered by some to have invalidated the notion of providence, due to an incommensurability with simple linear determinism. In an era when quantum physics has made the mechanistic paradigm redundant, a comprehensive review is urgently needed of all the (psychological, sociological and physical) factors at work in the structures and apparent anomalies of religious doctrine and belief. A stubborn unwillingness to consider the possibility that certain models of a providential process may be supported by certain legitimate interpretations of physics is hardly in the tradition of disinterested scientific enquiry.

Such a research strategy is obviously beyond the competence of any individual discipline within the natural or social sciences, and perhaps the abdication of theology as the 'Queen of the Sciences' was premature after all. In this regard, and in the light of the Waco disaster, the endowment of a lectureship in theology and natural science at the University of Cambridge (see *Nature* 362, 380 & 689-690; 1993) is to be welcomed by sceptical scientist and committed believer alike.

Michael Houlder
37 Argo Street,
Bolton BL3 6TT, UK

SIR — Just as science has its literature so does religion. The literature of science is concerned with observed facts of nature and with theories that attempt to encompass as many empirical observations as possible. The literature of religion is based mainly on interpretations of the Bible for the Judaeo-Christian religions and of the sacred works of the other religions of the world.

While the earliest attempts at explanation of the Universe were based on the thinking of the writers of the Bible and biblical scholars leading to the story of the creation in Genesis, other more scientific theories were being pronounced by Aristotle and other Greek philosophers.

However, by the end of the seventeenth century, most of this 'biblical science' had been replaced by the observations of Tycho Brahe and their interpretation by Kepler, and Newton in the fields of astronomy and mathematics. From this time onwards, science was founded on a firm empirical base while religious interpretations continued to be made on the basis of

faith in sacred writings.

The existence of God is a question that has been answered in numerous ways by religious philosophers and by scientists.

The main fact that has been established in Christian religion is that the man Jesus Christ did actually exist. It is a well documented matter of history that he had an exemplary life and that his teachings formed a valuable code of moral behaviour.

A belief in God requires a step in faith which cannot be justified by pure logic. It seems that God has an existence only in the minds of men who have taken this step and in their writings. These men have created a God in their own image with no historical or scientific basis.

It is to be hoped that whoever is elected to the lectureship in science and theology at the University of Cambridge will bear these facts in mind.

D. H. Evans
School of Science,
Sheffield Hallam University,
Pond Street, Sheffield S1 1WB, UK

SIR — Science has its base in natural philosophy. The scope of science, by virtue of its method of study, is limited to the 'sensible' Universe. Any exploration beyond that which can be directly or indirectly measured (even if only in principle) must take with it an element of *a priori* philosophy which cannot, in principle, be established scientifically. Outside the realm of the sensible, the scientific method of analysis has little to say. Scientists may, if they wish, postulate that there is nothing outside the sensible realm, but such theories are unprovable and, therefore, bad science.

Theology, on the other hand, explicitly allows for a God or gods whose being, partly or wholly, lies outside the scope of the sensible Universe (for example by virtue of being its Creator). Given this postulate, it is entirely reasonable to bring to bear philosophical reasoning that is not bound by the constraints of science.

It is true that there are many in both camps who vehemently contradict the views of the other, but I, as a scientist and a Christian, have my science and my religion, provided I am not required to apply scientific analysis beyond the bounds of its philosophic legitimacy. All serious thinkers should take this as a challenge to offer up examples of such obstacles.

Graham L. Mathieson
24 Courtmount Grove
Cosham, Hampshire PO6 2BL, UK

SIR — Josephson requests more scientific analysis of religion, but suggests no experiments or testable hypotheses.

There are two universal functions of religion. The most primitive is to explain the capricious nature of random events.

When humans first became aware of the future and started planning, they also realized that the cause of some events was not visible. The reasonable deduction was that some more powerful being or beings decided the weather, the success of hunt and harvest, the casting of lots. This role is important beyond the simplest human societies. The statistics of probability are relatively new discoveries, and nearly all children and about half the educated adults of the industrialized world use absolute reasoning, where every effect has an exact cause, and the cold virus comes literally from cold draught and damp ground.

The second role of religion is in extended societies. Humans presumably evolved in small family troupes with a dominant individual in command. As societies grew, a unifying authority was needed to bolster the figureheads whose human failings became obvious. The need of humans for a leader is clear in the attention given to royalty, presidents, prime ministers and dictators, especially during wars. Their validity has been bolstered by divine right and papal infallibility. Islam unites state and religion. A higher ideal, be it God, communism or a welfare state, is needed rationally to justify the self-sacrificing co-operation needed for stable agricultural and urbanized societies where the bounds of kin selection are exceeded and extended. Without such unifying aims, there is no objection to giving selfishness, genetic or otherwise, reign in a Clockwork Orange society. The natural goodness of humanity is quite apparent in Bosnia, where three societies meet, each with their own religion. The only 'scientific' analysis of religion seems to be subjective interpretation through the hindsight of history: how did religion or its absence affect development of societies? Humanist society in a densely populated world puts selfish genes in direct conflict with genes for mutual co-operation. I pessimistically expect the former to prosper in a frequency dependent manner. The questions that history will answer are these. Does human mass-psychology absolutely need a God? Can a stable, cohesive society persist without the carrot-and-stick glue of religion?

Hugh Fletcher

Queen's University of Belfast,
Medical Biology Centre,
97 Lisburn Road,
Belfast BT9 7BL, UK

SIR — Josephson's suggestion that "religious practices have in part a genetic basis" may sound vague but reasonable. Suggesting that "a central theme of religion is the attempt to maximize human 'goodness'" and that "genes linked to the potential of goodness" are involved is not even reasonable. Religious practices are causing bloodshed that masks any attempt

to describe them as a way to "maximize human goodness".

The description of these awful consequences of religions as "pathological variants" is an attempt to put practice aside and to focus on the ideology and on the original purposes of religions. However, it is impossible to obtain first-hand witness to the ideology and the original purposes of any religion. We can only study the written ideologies (holy scriptures) as well as their actual practice over history in an attempt to uncover their purposes. There is unlikely to be agreement as to these purposes.

Instead, therefore, of defining religion in a questionable way and then investigating religion scientifically, it might be appropriate to leave religion out of the biological sciences and to focus on better defined traits that might have a selection advantage for individuals who carry them, by helping those individuals to be part of a society. Such traits could indirectly increase 'human goodness'.

David Granot

Institute of Field and Garden Crops,
Agricultural Research Organization,
The Volcani Center,
PO Box 6, Bet Dagan 50250, Israel

SIR — The concern expressed in the British parliament about people who join 'sects' poses a question. What is the difference between a religion (for example that headed by the Pope), and a sect (that led by David Koresh which met its tragic end in Waco)? The answer seems to be that if it is old-established and has a large number of followers, it's a religion (and respectable), whereas if it is new and followed only by a few, it's a sect (and therefore not respectable). For those of us who are not blinded by 'faith', it is clear that any differences are only of degree, not of substance. So, what is this madness which is religion all about?

Josephson proposes a genetic basis for religious belief, and speculates that, as the theme of religion is the attempt to maximize human "goodness", societies holding such beliefs will function "more harmoniously and efficiently, so that natural selection will tend to favour the presence in human societies of genes of this type". I think most scientists would accept the former premise — that religious belief has a genetic basis — but not the latter, that such beliefs have anything to do with 'goodness'.

It is important to emphasize that it is the 'need' to 'believe' that is genetically determined, and not the beliefs *per se*, which are obviously culturally based. The notion that religious belief is genetically determined is not new. E. O. Wilson, in his book *On Human Nature* (Harvard University Press, 1978), argues (page 188) that "religious practice . . . can be seen to confer biological advantage". This,

however, has nothing to do with 'goodness', but provides each believer with "unquestioned membership in a group". This then appears to provide the justification for almost unlimited viciousness to outgroups, as current events around the world amply demonstrate. Wilson sees a ray of hope. He writes that "the final decisive edge enjoyed by scientific naturalism will come from its capacity to explain traditional religion, its chief competitor, as a wholly material phenomenon. Theology is not likely to survive as an independent intellectual discipline." One can only hope that the Starbridge lectureship will hasten this, unlikely though that seems.

There are other grounds for optimism. The human genome project will sooner or later identify the gene or genes responsible for religious belief. Perhaps we can contemplate (with tongue only slightly in cheek!) a brave new world in which genetic engineering can free humanity from the scourge of religion and allow us to look forward to a bright rationalist future.

Christopher J. Lote

Physiology Department,
University of Birmingham,
Birmingham B15 2TT, UK

SIR — I want to ask only one question of the advocates of a lectureship in theology and natural science in the University of Cambridge: What have the teachings of Jesus Christ, as recounted in the Bible, to do with the complexities and intricacies of molecular biology? There is a close relationship between religion and morality and ethics — although many atheists are righteous and kind-hearted even without the fear that a higher being may punish their wrong-doings — but there is no connection at all between theology and natural science *per se*.

That great scientists were believing Christians does not prove anything. In this century, many free-thinkers have also made great contributions to science, scientific thinking and ethical questions regarding the applications of science.

Friedrich Katscher

Mariahilfer Straase 133,
A-1150 Vienna, Austria

Correction

In a letter headed "Religion in the genes" by Brian D. Josephson (*Nature* 362, 583; 1993), the words "non-experiential" and "experientially" were inadvertently changed to "non-experimental" and "experimentally".

The end of the third paragraph should read: "([Humanism] shares the general goals of religion, but adopts a primarily non-experiential approach towards their achievement.)" In the fourth paragraph, the second sentence should read: "Some [religious cultures], for example, take divinity as a central theme (experientially as well as intellectually) . . .".

Reference 1 should have read "Sutherland, S. 361, 292 (1993)". □

AIDS: striking the happy media

John Moore

The past year has seen many controversies about AIDS research and researchers. What productive events have occurred, and what is likely to happen in the next year?

IN 1961, John F. Kennedy stood at a podium in Berlin and announced "*Ich bin ein Berliner!*" Colloquially, this can mean "I am a doughnut". Will many AIDS researchers do the same at the IX International AIDS Conference in Berlin next week? The international AIDS meeting has long since shot its bolt as a worthwhile forum for scientific debate — it is far too large, unfocused and glitzy for many scientists' tastes. But it does attract the media, on the look-out for stories. Last year's Amsterdam meeting was marred by the extraordinary focus on what transpired to be a minor issue: HIV-negative AIDS, or ICL. A premature announcement created unnecessary public anxiety about a possible new epidemic. All the ICL scare did was to attract attention away from the real issues: the plight of people infected with HIV and the attempts of science to help.

AIDS, with its combination of sex, death and celebrities, holds a strong fascination for the media. Thus there is an onus on scientists to deal responsibly with journalists. It is a common perception in the scientific community that inappropriate press coverage of AIDS issues is the fault of journalists. Except for some tabloids like the London *Sunday Times*, this is usually not the case. Journalists working for specialist science magazines and most quality newspapers are generally responsible professionals who report the field fairly and accurately. Journalists earn their corn by writing stories, but often the bread is given to them already buttered. By and large, press coverage of AIDS issues reflects what scientists say to journalists. A journalist's responsibility is to check that the facts are accurate, but not necessarily to judge their overall merit. Why should a good story be spiked just because other scientists disagree with the data interpretation? When scientists say contradictory things to the public, how can the public assess whom to believe?

Dissent and controversy

Science has a duty to inform and educate the public, but it must neither frighten people unnecessarily nor give them unjustified expectations. Claims of "AIDS cures" in the popular press need to be based on much more than just *in vitro* data. Whatever the need to attract research funding, is 15 minutes of fame for one person ever worth 15 days of fear or 15 weeks of false hopes for many? Another

worrisome trend, especially in the United Kingdom, is the tendency of some AIDS researchers to whine to the press and politicians about their failure to win competitive funding, alleging a non-existent bias by funding organizations. These antics create an awful impression on the taxpayers and politicians who pay for our work, and obscure the real message: more resources need to be committed in a more effective fashion to fight against AIDS.

The consequences of dissent and controversy are now apparent, at least in the United Kingdom, where hostile press comment may have contributed to the recent change in government AIDS policies. This will be to the detriment of those infected with HIV and those yet to be infected, as well as everyone involved in AIDS research and education. The 'Murdoch' press has run a highly slanted campaign parrotting Peter Duesberg's line that HIV is not the cause of AIDS and that the risk to the heterosexual population is minimal. The subliminal message of these newspapers is that "normal" people (*Sunday Times* readers?) don't get AIDS. There is dirty politics at work, but have we scientists contributed to the current climate by our interactions with the press? Already there are signs of a backlash against AIDS issues in the US press. I believe it is the responsibility of all of us engaged in AIDS research to take a broad view when dealing with the media: hyping our own laboratories' achievements or boosting our companies' share-price may not be in the wider interest of AIDS research.

As well as the ICL fiasco, there have been several major AIDS stories since the last AIDS conference. Their significance to science has usually been a lot less than their perceived importance in the press, but can we learn any lessons from them? Duesberg's anti-HIV polemics continue to attract press attention, but to most specialists the case for HIV is proved beyond reasonable doubt and the arguments have become sterile. Duesberg reminds me of the Black Knight in "Monty Python and the Holy Grail". His limbs cut off one by one in a duel, the torso waddles towards his antagonist and threatens to bite his kneecaps! Sadly, this is no laughing matter — Duesberg's nonsense can and has hurt people.

This year there has also been a debate in the press about whether HIV could have been introduced into Africa in contamin-

ated poliovirus vaccines. The hypothesis was shown to be highly improbable, and although it would be useful to know where HIV came from, it seems to me more important to know where it is going and to stop it getting there. Similarly, the press fascination with who first isolated HIV-1 is becoming a bore, and worse. The witch-hunts just stop people from getting on their work. Now it is clear that many people played crucial roles in the isolation of HIV-1 (J. C. Gluckman, *Science* **259**, 1809; 1993), perhaps it is time to let sleeping dogs LAI!

Media manipulation

The failure of AZT to evoke long-lasting clinical benefit attracted much comment when the Concorde trial results were published recently (*Lancet* **341**, 889; 1993). These were important findings. But what was news to the public was not to most AIDS researchers; the development of AZT-resistant variants *in vivo* is well known. The strategy for anti-retroviral drug therapy has been focused on drug combinations for some time now. A publication describing *in vitro* effects of three antiviral agents used in concert (M. Chow *et al.* *Nature* **361**, 650–653; 1993) attracted far more press coverage than was warranted by its scientific content. The publicity engendered by this paper seemed to me not to be in the best interests of the public or of AIDS research. If we cry wolf too often, people will stop listening when there is something really important to say.

Attempts by the MicroGeneSys company to acquire 'pork-barrel' funding from the US Congress for its gp160 'AIDS vaccine' by slick use of the lobbying system were defeated this year by a triple combination of accurate reporting in the press, scientific pressure and the efforts of AIDS activists who knew a con-trick when they saw one. In short, a triumph for science over the profit motive. The fuss originated from claims by US Army scientists at last year's Amsterdam conference that immunization of HIV-infected people with gp160 reduces viral burden and stabilizes CD4 counts, events thought to be of clinical benefit. These claims require confirmation and are still the focus of an official investigation. But the hopeful climate created enabled the manufacturers of gp160 to sneak a \$20 million appropriation onto the Department of Defense finance bill. This sum could have

supported about 40 RO1 investigator grants for 5 years. A successful campaign against this mode of science funding resulted in the money being transferred to the NIH for comparative therapeutic trials of different candidate vaccines, the best outcome that could have been achieved in the circumstances. As the NIH, unlike the US Army, have refused to pay any company for their products, it appears that the MicroGeneSys' 'vaccine' will not be in the comparative trial. The company's expenditure on lobbyists could ultimately benefit only its rival companies — a truly appropriate outcome.

One important issue came out of the MicroGeneSys battle: should efficacy standards of clinical trials be lowered when considering potential AIDS therapies? This is a very tough call. An NIH/FDA advisory panel voted by 15 to 1 that there were no traditional scientific reasons to undertake a phase III trial of any therapeutic vaccine candidate, but then voted by 16 to 0 to undertake a trial essentially on compassionate grounds. The decision was understandable in the social and political climate in which it was taken, but it does set a dangerous precedent that companies with products they view as promising will no doubt exploit.

The US Army has perhaps taken this approach a step too far. Its advisory group decided that there was no "requirement for evidence of *likelihood* of efficacy" (my emphasis) for a product to get into a US Army clinical trial. Would the US Army equip its tanks with guns that had no likelihood of firing? The problem with therapeutic immunization is that while few AIDS researchers think it will work, nobody knows for sure that it won't. Thus it is hard to argue against the emotional argument that immunization might save the lives of people dying of AIDS. Ethical dilemmas abound in AIDS research, but overall a controlled trial of the concept is warranted. Yet FDA approval of a therapeutic vaccine may well require clinical endpoints of greater validity than surrogate markers.

Irrespective of how the media reports the Berlin conference, what important things might happen in selected areas of AIDS research over the next year? There is a rapidly emerging consensus that AIDS is a virus infection and not an autoimmune disease. Recent studies from several groups quantitating viraemia in both the peripheral blood and lymphoid tissues, and showing consistent changes in virus load and CD4⁺ counts in response to antiviral treatment, should be sufficient to quiet the doubting Thomases. Although there will still be attention focused on issues such as apparent gp120 sequence homologies to MHC class II and mycoplasma adhesion sites, on cross-reactive antibodies, superantigens, and bizarre epi-phenomena in autoimmune mice, in-

creasingly they will be seen as peripheral to the central verity that HIV is a cytopathic virus with CD4⁺ cell tropism.

We should also see an increasing awareness of the importance of cellular immunity as a limiting factor to the spread of viral infections. This was clear to some people years ago, but is only gradually sinking into the consciousness of many of us. Behind this lack of appreciation has been the absence of any consensus among immunologists about how HIV cripples the immune system. As an extreme example, most cytotoxic T lymphocytes are CD8⁺ cells, which are seen by many as important in fighting viral infections. Yet others believe with equal passion that we should be trying to eliminate them.

Among emerging technologies that may eventually reach the clinic are vaccines based on immunization not with proteins but with DNA. Although counter-intuitive, this approach does seem promising based on animal experiments. As we understand more about AIDS pathogenesis, the human immune system and vaccine design, it is becoming more likely that the correlates of protective immunity encompass both humoral and cellular mechanisms. If that sounds like a reinvention of the wheel, that's almost a tradition in AIDS research. But it seems reasonable to assume that the success of Desrosiers' attenuated, live-virus vaccine in the SIV model is because it stimulates continuous-ly all arms of the immune system.

The recent trend towards cooperation between companies working on antiviral drugs is both constructive and refreshing. The concept could be extended profitably to vaccine development. Subunit vaccines currently under evaluation tend to have been designed several years ago on an *ad hoc* basis but still must be tested rigorously. Even should they fail, we may learn much of help in designing the next generation of reagents. Different products have different useful features. For example, Genentech, Chiron and others have made gp120 vaccines that are effective stimulators of humoral and cellular immune responses to gp120, but which lack *gag* components against which some cellular immunity is directed. The 'Salk vaccine' was accidentally depleted of gp120 in the production process. Thus while it might be a good stimulator of cellular immunity, it lacks the ability to evoke neutralizing antibodies to the viral envelope. Current theories are sufficiently controversial for it to be uncertain whether this is a benefit or a defect. The companies could consider joining forces to compare the effects of their products in isolation and together. An analogous situation can be foreseen with therapeutic antibodies, where different companies own antibodies that might work best in combination. There are lives at stake as well as money, so would anybody be

happy if several companies have half an AIDS drug or vaccine and the world has none?

A major development in AIDS research in 1993 will be organizational, with the creation of the Office of AIDS Research (OAR) as a coordinating council for US government-funded research programmes. It seems ironic that as the UK AIDS Directed Programme comes under pressure from the Medical Research Council, the Americans are copying its steering committee format. It is difficult to predict how the OAR will affect current trends and practices in AIDS administration, as the organization so far lacks a director(s) and an advisory board. Many scientists fear the OAR, through ignorance of its role; others feel it may merely add an unnecessary layer of bureaucracy to AIDS administration; some are concerned that funding priorities will be decided by AIDS activists, who undoubtedly wield more power under Clinton than under Reagan or Bush. I believe these fears are groundless. The role of activists in the OAR should be viewed optimistically, as its planners intended. The role of activists in the OAR is positive; the current generation of activist leaders is a responsible one with much to offer the scientists who are prepared to work with them in partnership. Nevertheless, scientists should ultimately decide about scientific issues. The OAR will establish funding priorities and focus resources on key research areas as they emerge; it should not micro-manage research in individual research labs and it should not support administrators at the expense of scientists.

It is to be hoped that the OAR will acquire new funds for AIDS research, but equally important is to spend existing funds more wisely. There is waste and unnecessary duplication in AIDS research, and some work that is poor. All increases in scientific knowledge are incremental, but in AIDS research sorting the increment from the excrement is a real problem. Yet statistical analysis shows that AIDS literature is not much different from other areas of biomedical research (P. Brown, *New Scient.* 15 May 1993). Large sums of money have been thrown at AIDS research; merely scattering more around is not the answer. It must be well-aimed to be effective.

The younger generation of AIDS researchers tends to cooperate closely, not just to survive in a cut-throat world, but out of the realization that AIDS is bigger than all of us and that it will need a meld of all our skills to defeat it. As JFK might have put it: "Think not what AIDS research can do for you; think what you can do for AIDS research". □

John Moore washes ELISA plates at the Aaron Diamond AIDS Research Center, New York, New York 10016, USA.

AIDS: trends, predictions, controversy

Roy Anderson

One of the persistent controversies that surfaces in the media about AIDS is whether the heterosexual population is at risk. The latest projections provide an emphatic affirmative.

PEOPLE working in AIDS research may be feeling somewhat besieged at present with frequent press comment on controversies such as the potential magnitude of the epidemic in developed countries and on the frustratingly slow progress in drug and vaccine development. The lack of balance in reports in some sectors of the press and media has not helped, with a particular trend in recent coverage of AIDS being the assertion that epidemiologists have greatly exaggerated the potential of the virus to spread in heterosexual populations in developed countries, and even a denial that there is a major HIV epidemic in sub-Saharan Africa despite, in the latter case, numerous studies that chart the steady rise of infection in low- and high-risk groups in urban centres and increasingly in rural areas in these regions¹. The extreme views presented by some, in particular the *Sunday Times* in giving generous space to the Duesberg sect, have been balanced by careful and accurate reporting by others, notably the *New York Times*. But the diversity of views presented has created confusion in the minds of the public, researchers in other fields, and policy makers.

What is the current pattern of spread of AIDS in industrialized countries? Many of them have sophisticated surveillance systems that reveal the temporal evolution in the incidence of AIDS within different at-risk groups. The median incubation period of AIDS (time from infection to the development of AIDS) appears to be approximately 10 years in sexually active adults in developed countries irrespective of risk group, hence reports of the current incidence of AIDS reveal the pattern of HIV spread a decade ago. To improve the information base available to planners and policy makers, many countries have developed unlinked anonymous monitoring of HIV prevalence among people attending medical clinics².

In Europe, temporal trends in the incidence of AIDS in most countries depict a steady rise over the past decade (see Fig. 1)³. But there are important differences between countries in the overall magnitude of the epidemic and its constituent pattern in different at-risk groups. For example, cumulative AIDS cases per million head of population at the end of 1992

were approximately 1,200 in the United States, 441 in Spain, 417 in Switzerland, 403 in France, 272 in Italy, 120 in the United Kingdom and 89 in Sweden³. For the last year for which accurate data are available (1991–92) in Europe as a whole the largest percentage increase occurred in the heterosexual transmission group (24.8), although there is much variability between countries. In the United States and the United Kingdom, cases in homosexual and bisexual men form about 60–70% of the total, but cases among heterosexuals now exceed those among intravenous drug users to form the

that can tentatively be drawn from these observed patterns. The epidemic is still in its early stages in most risk groups, although there are encouraging signs of a slowing in the rate of accumulation of new AIDS case reports in homosexual/bisexual men in most developed countries. The long duration of the epidemic is a simple consequence of the long average incubation period of AIDS. Its full extent will not be known for many decades. This point has been made numerous times over the past 7 years³, but the message of a slowly developing epidemic, particularly in the lower risk groups (heterosexuals), is

still not accurately reported in many sectors of the press. To assess the likelihood of a more widely disseminated epidemic among heterosexuals, the important point is to base this assessment not on the comparatively low numbers of AIDS cases in this group at present, but on the trends in AIDS incidence and HIV seroprevalence over the past decade. These reveal an expanding epidemic, moving on a very slow timescale, largely centred at present in urban centres in poor communities and in socially marginalized groups.

There remain many uncertainties about the key epidemiological determinants of the epidemic. These include the likely pattern of the full incubation period distribution of AIDS; the fraction of those infected who will eventually develop AIDS; fluctuations in infectiousness during the incubation period; the relevance to transmission and pathogenesis of the genetic diversity of the virus; the prevailing heterogeneities in, and frequencies of, the behaviours that determine transmission; patterns of mixing or contact within and between the different at-risk groups; and the influence of education on these behaviours and mixing patterns. These uncertainties will not be resolved quickly.

Projections of future trends are obviously required, despite these uncertainties, to plan sensibly for health-care needs and to assess priorities in expenditure. Projections can only be made in quantitative terms for the short term (2–4 years). The methods available include extrapolation on the basis of simple mathematical functions that are assumed to mirror the future shape of the

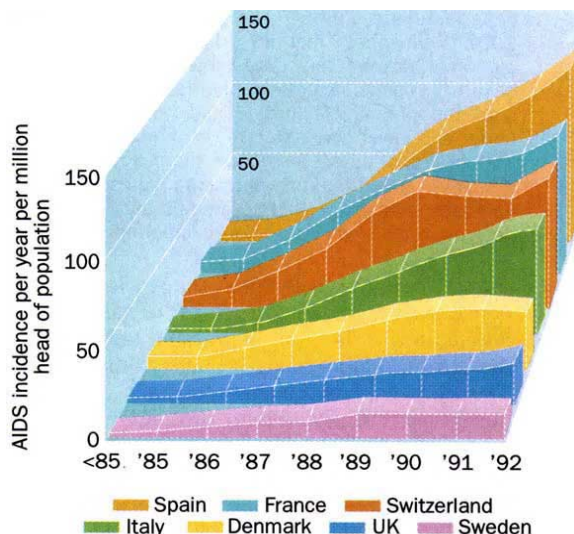


FIG. 1 Reported incidence of AIDS per annum, corrected for reporting delay, in a selection of European countries (data from ref. 13).

second-highest risk category. Of the new reports of AIDS in the United Kingdom during March 1993, for example, 60% were in homosexual or bisexual men, 15% in heterosexuals and 9% in injecting drug users. Conversely, in southern European countries such as Italy and Spain, injecting drug use is the major risk. Even within countries, however, the pattern varies². Trends in HIV seroprevalence are more difficult to obtain but reports from a European study⁴ of patients attending sexually transmitted disease clinics revealed wide variation. Among non-intravenous drug using heterosexuals, prevalence ranged from 1–2% in France, Germany, Italy, Spain and Switzerland, to 0–0.3% in the United Kingdom, Sweden and Denmark.

There are several general conclusions

edipemic⁵, complex mathematical models that attempt to mimic the transmission dynamics of the virus between and within risk groups⁶, and what are called back-calculation methods⁷. Extrapolation based on simple functions is unreliable due to the non-linear and uncertain shape of the epidemic. Transmission models contain many parameters, particularly with respect to sexual behaviour and mixing patterns, for which estimates are not available at present; hence the favoured method is back-calculation.

The principle behind it is simple. Given information on the incidence of AIDS from surveillance reports plus the probability distribution of the incubation period estimated from cohort studies, it is possible to back-calculate the number of people infected with HIV during each time period necessary to account for the observed AIDS case reports. The number of people who will be diagnosed with AIDS in the near future (2–4 years) is then projected by extrapolation based on the shape of the estimated HIV incidence curve and the incubation period distribution. Most countries rely on this method, and reports in the United States⁸ and the United Kingdom⁹ have, in general, produced reasonably reliable projections when measured against the observed outcomes. More recent calculations much more closely match observed trends (Fig. 2) whereas the early reports (1986, 1988) overestimated the degree of spread. The greatest uncertainty in future trends has been, and still is, among heterosexuals.

The main reason for the downwards revision in the projected incidence of AIDS in the coming few years lies in the trend revealed by the back-calculation method in the incidence of new HIV infections among homosexual and bisexual men over the past decade, where a sharp slowing down in the rate of transmission appears to have occurred from about 1984. Interestingly, this trend is revealed in most developed countries and in both of the two major risk groups (homosexual men and intravenous drug users). By contrast, estimated incidence curves for heterosexuals show an upward trend over the past decade (Fig. 2).

Three factors, or some combination of them, could explain the trend in homosexual men. First, it could simply reflect changes in behaviour as a result of education and media coverage. Second, it could be a reflection of the natural pattern of the epidemic, where saturation of infection in the high-risk groups reduces the rate at which secondary cases arise. Third, it could be spurious, and a direct consequence of the assumptions underlying the back-calculation method. The first explanation is probably an important component, but direct quantitative evidence of behaviour change is limited. The second explanation must also be important, as

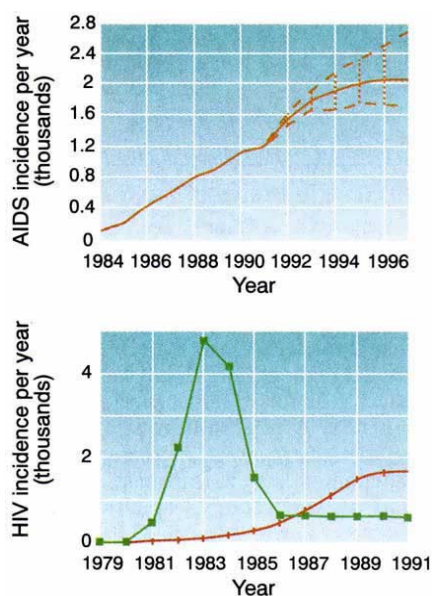


FIG. 2 Top, latest projections of AIDS incidence per annum in England and Wales in all risk groups combined⁸. Estimated HIV prevalence at end of 1991: lower estimate, 20,000; upper estimate, 29,050. The solid line up to 1992 is essentially identical to the observed outcome. Beyond this date projections are based on back-calculation methods and record the planning figures (solid line) and upper and lower bounds on the projections⁹. Bottom graph, estimated annual incidence of new HIV-1 infection in England and Wales in homosexual/bisexual men (green) and in heterosexuals (red) from back-calculation⁹.

simple theory reveals a bell-shaped incidence curve for infections in a naive population, with a long right-hand tail in the absence of any change in behaviour¹⁰. The third explanation has not been explored sufficiently. First, the detailed back-projections of HIV incidence (see Fig. 2) are sensitive to the assumption of when the epidemic started. Second, the method tends to allocate most HIV infections to as early a time point as possible. Third, the full incubation period of AIDS and the fraction of those infected who will develop AIDS are unknown¹¹. Last, the method is not suitable for assessing trends in incidence in the few years immediately preceding the last observation on AIDS incidence (the flat segment on the right-hand side of the bottom graph for homosexual men in Fig. 2). Put together, these problems imply that the estimated magnitude of the HIV incidence curve may be too great early in the epidemic and too small in the most recent time frame. These methodological issues highlight the uncertainty surrounding future trends.

Simple and complex models of the transmission dynamics of HIV based on current data of the key processes (admittedly limited) suggest that, in the absence of a major breakthrough in research on drugs that halt or slow progression to AIDS as well as reducing the infectiousness of infected individuals, and without

hope of vaccines to prevent infection, HIV is likely to attain an endemic state in the coming decades. The epidemic pattern before this state may be multi-peaked, perhaps with each peak separated by many years as infection passes from high- to medium- to lower-risk groups. The magnitude of the endemic stage, and the time taken to attain it, will vary between countries, as do the prevalences of other sexually transmitted infections. It appears likely that the endemic state in Europe and North America will be maintained largely within the core groups of individuals with high-risk behaviour belonging to each of the major at-risk categories. For sexually transmitted diseases in general the size of the core group varies according to the biological properties of the infection and the behavioural determinants of transmission. Endemic prevalences of other sexually transmitted diseases are not necessarily good guides to the likely extent of HIV spread in future decades. However, taking genital herpes as an example, surveys in London reveal that roughly 7–10% of pregnant women of Caucasian origin aged 35–40 have evidence of past infection¹². There are obvious dangers in drawing parallels between these viruses; but these studies highlight what may be an upper bound to the potential size of the heterosexual group at risk.

The relative effectiveness of AIDS awareness and sex-education programmes for young people before they start sexual activity will obviously be important. Given the many uncertainties, a precautionary approach seems advisable, with education targeted broadly, as no one can predict to which, if any, risk group, a young teenager will eventually belong. AIDS is a dreadful disease which is invariably lethal, given current treatment options. The media therefore has a responsibility to report the facts correctly and to avoid sensationalism, whether through exaggeration or by denial. □

Roy Anderson is in the Biology Department, Imperial College of Science, Technology and Medicine, London SW7 2BB, UK.

- Chin, J., Remenyi, M.-A., Morrison, F. & Bulatoa, R. *World Hlth Stat. Quart.* **45**, 22–27 (1992).
- Commun. Dis. Rep.* **3**, R1–R16 (1993).
- Anderson, R. M. & May, R. M. *Nature* **333**, 514–519 (1988).
- Morbidity Wkly Rep.* **41**, No. RR-18, 1–29 (1993).
- Norman, C. *Science* **230**, 1020–1 (1985).
- Hyman, J. M. & Stanley, E. A. *Math. Biosci.* **10**, 415–473 (1988).
- Brookmeyer, R. & Gail, M. J. H. *J. Am. Stat. Ass.* **83**, 301–308 (1988).
- Aboulker, J. P. & Short, A. M. *Lancet* **341**, 889–90 (1993).
- PHLS AIDS Centre at CDSC. AIDS Projections to 1997, England and Wales. *Comm. Dis. Rep.* (in the press).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans* (Oxford University Press, 1991).
- Hendriks, J. C. M. *et al.* *AIDS* **7**, 231–239 (1993).
- Ades, A. E. *et al.* *J. Epidemiol. Comm. Hlth* **43**, 53–60 (1989).
- WHO-EC Collaborating Centre on AIDS *Quart. Rep. No.* 36, 1–39 (1993).

Calculating the energy of fullerenes

So far, the experimentalists have had the most fun from fullerenes, but now the theorists are catching up. Not before time, for they have to live down the shame of not having predicted them.

THAT the discovery of such molecules as C_{60} has been a surprise goes without saying. That they should be made so crudely, for example by collecting the soot from a carbon arc and not by a team effort in some organic chemistry synthesis laboratory, is only part of the tale. That truncated icosahedral cages of carbon atoms should be stable relative to, say, flattened sheets is another. That they should also be chemically akin to alkenes more than to benzene is yet another. And since the discovery of the simplest fullerenes, larger cage structure made only of carbon have been found, as have been the famous nanotubes which may yet become (when filled with suitable metallic atoms) the first electrically conducting nano-wires and the still more puzzling structures consisting of a small number of concentric shells of carbon atoms held together as if there were an onion. Sceptics must be forgiven for asking where it will all end.

Inevitably, the search for new forms of this unexpected allotrope of carbon has been matched in intensity only by the attempts of theorists to calculate the laboratory features of the fullerenes so far observed. The simplest tasks are easily done. Electronically, C_{60} is the three-dimensional analogue of benzene, so that the electron states can be calculated by a simple extension of material found in many text-books. The thermodynamic properties of the crystalline form, and in particular the degree to which rotation of the cages is reduced with decreasing temperature, can be estimated, most accurately by molecular dynamics. It remains a puzzle that interstitial potassium ions appear to turn the C_{60} solid into a superconductor, but even the superconductivity of the copper-oxide ceramics remains obscure except to those who must be supposed to understand their own theories.

Similar tasks have been carried out for larger fullerene molecules, if and when there is some guidance about the symmetry of the molecular structure (or the molecule is small enough for the alternatives to be enumerated). Meanwhile, it has ceased to be a cause for wonder that the supposed baseline of the energetic stability of carbon atoms in bulk — the electronically conjugated carbon atoms in a sheet of graphite — so readily curls up into icosahedral molecules. An infinite sheet of carbon atoms may be the most stable form there can be, but a finite piece of the same structure has so many unrequited or "dangling" chemical bonds that the energy of bond-bending when forming even a

C_{60} cage is far outdone by the energy of forming the covalent bonds that hold the structure together.

In passing, it is not irreverent to recall that the existence of C_{60} was almost entirely overlooked by theorists until the molecule showed up in carbon soot. For the past quarter of a century, people have been working with conjugated electron systems whose complexity has steadily increased. Yet the suggestions that a spheroidal form of graphite might be stable were few and little heeded before its discovery in 1985. E. Osawa proposed in 1970 that C_{60} might be stable, and the Soviets D.A. Bochvar and E.G. Gal'pern went so far as to perform molecular-orbital calculations in 1973; while at the University of California at Los Angeles, Orville Chapman attempted the very synthesis alluded to earlier. But somehow, the attractive idea of an icosahedral cage molecule failed to spread wider roots.

Whatever the explanation, and despite the variety of molecularly distinct forms (isomers included) of the fullerenes found so far, there seems still to be a long way to go before this surprising ode is worked out. That at least is the inference from the work of J. Bernholc and two colleagues at the University of North Carolina (Maiti, A., Brabec, C. & Bernholc, *J. Phys. Rev. Lett.* **70**, 3023; 17 May 1993). Although there has been a spate of attempts to calculate the energy of particular fullerenes in the past four years, Bernholc and his colleagues reach some startling conclusions about fullerenes not yet made by a refreshing mixture of verbal argument with calculation. (More formally, they call it "scaling arguments and explicit energetic considerations".) The calculation part consists of molecular dynamics in which spurious energy minima (corresponding to fullerene cages with defects) are avoided by imposing an artificial symmetry on the system.

Perhaps the most startling result is not a result at all, but a kind of laboratory notebook entry; the calculations have been extended to cage-like fullerenes with more than 6,000 identical carbon atoms in order that their energy can be compared with that of the onion-structured isomers, with a few concentric cages rather than one large cage. The surprising result is that the most stable (or least energetic form) of the single-shelled structure is always a polyhedron (so that puckered shells are unfavoured aberrations). Among isomers for a fullerene of specified even-integer rank, the most stable form is always that in which there are exactly twelve

pentagons. (Fullerenes of odd rank are intrinsically odd; they cannot satisfy the condition that there should be exactly three bonds at each vertex.) Icosahedral symmetry is also an advantage.

Otherwise, the outcome of the calculations is so regular that the authors are able to fit it by a graph for the energy per carbon atom in the shell which is a smoothly declining function of the number of atoms and which appears to be accurate to a few parts in 10,000 for shells containing up to 10,000 carbon atoms. The snag is simply that the energy of the lowest possible state of a shell of specified size may not be very different from others only slightly less stable. That is why the special symmetry of the theoretical ground state may not be that found in crystals of these super-fullerene molecules — if the materials themselves are ever found.

Armed with information, the business of calculating the onion-structures is then straightforward. It is merely necessary to arrange that all shells will indeed fit within the next largest as if it were a Russian doll slipped inside the next largest in the sequence. But the onion structure become stable relative to the single shell only when the number of atoms is so great that the mutual attractive energy of the concentric layers exceeds that associated with the extra curvature of the surfaces.

None of this is earthshaking. On the contrary (and apart from the rule that the most stable full recognizable shape must be an open question, especially if there is so little difference between the energy of the ground state and those next to it in stability).

In any case, there is plenty of work on fullerenes to occupy the theorists for some time. The electronic structure of molecules such as these which appear to be deformable at low energetic cost seems an interesting problem. How the geometric constraints might influence the electronic band structure may be a still more subtle issue. Carbon nanotubes, for example, are predicted to transform from insulators into semiconductors merely by the variation of the helical twist in their chicken-wire structure. And with the further possibility of tuning the electronic and mechanical properties by stuffing these hollow structures with metals, or even by doping the carbon sheets themselves with atoms such as boron, the permutations available for theoretical exploration seem endless. (It seems unlikely that the theorists will be caught napping a second time.)

John Maddox

Siberian mice upset Mendel

Andrew Pomiankowski and Laurence D. Hurst

FACED with malaria, those who are heterozygous for the sickle-cell gene are immune to attack, and are at an advantage over both normal homozygotes and sickle-cell homozygotes (who suffer severely from sickle-cell disease). The pity then is that heterozygote parents cannot produce exclusively heterozygous zygotes in such circumstances, for a rule of genetics is that in such matings only half the progeny will have the parental genotype.

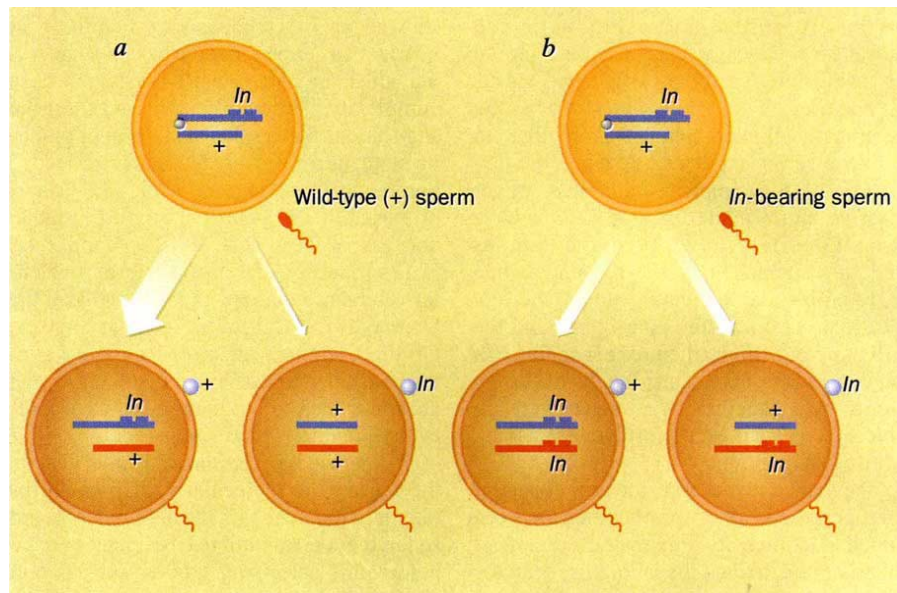
More broadly, going beyond that, it is a tenet of modern population genetics that meiotic segregation occurs independently of the fitness of the resulting zygote. Or does it? The answer it seems is 'no', at least in one instance, as the Russian team of Agulnik, Sabantsev and Ruvinsky show in a report just published in *Genetical Research*¹.

Agulnik and colleagues have been working, not on the sickle-cell trait, but on an aberrant form of chromosome 1 found in Siberian populations of house mice. The aberrant chromosome is about 30 per cent larger than the wild type because it contains two large, homogeneously staining insertions that are held in an inversion (termed *In*). *In* behaves strangely during oogenesis. If homozygous wild-type males (+/+) are mated to heterozygous females (*In*/+), about 85 per cent of the progeny inherit the insertions.

The transmission bias is the consequence of unusual segregation of *In* during oogenesis. Unlike spermatogenesis, in which all four of the products of meiosis are viable, oogenesis results in a single haploid gamete with the remaining three chromatids being discarded in polar bodies. Agulnik and colleagues have shown that, during the first and second meiotic divisions, the *In*-bearing chromatid often ends up in the egg and not the polar body². The alternative possibility, that transmission bias is due to differential mortality of +/+ zygotes, can be rejected as the genotypes of blastocysts and newborns are in close agreement.

This exception to normal mendelian segregation is an example of meiotic drive, a process in which a gene (or gene complex) manages to spread in a population by ensuring that meiosis is biased in its favour. Unlike the *In* system, the best studied examples of autosomal meiotic-drive genes, the *t*-complex of mice and *Segregation Distorter* (*SD*) of *Drosophila*, both have their effects on spermatogenesis³. In heterozygotes these driver-gene complexes disable wild-type sperm, thus ensuring that sperm containing the driver genes monopolize fertilizations by that male (assuming females are singly mated).

Despite the difference in mode of distortion, *In* has at least three similarities to *SD* and the *t*-complex. First, all of the gene complexes involve two types of loci, one that causes the distortion and another that confers insensitivity to distortion^{3,4}. Second, alleles at these two loci form a gene complex held in linkage disequilibrium by chromosomal inversion^{3,5}.



The *In* effect, in heterozygous oocytes just before the second meiotic division. The degree of segregation distortion induced by *In* is dependent upon the sperm that the oocyte has fused with. When fusing with wild-type sperm (a) about 85 per cent of progeny are heterozygotes. When fusing with *In*-bearing sperm (b) the transmission bias is annulled. The unused chromatid from this second meiotic division is thrown into a polar body (spherical blue object). The relative thickness of the arrows indicates the degree of transmission bias.

Third, drive homozygotes have reduced fitness^{2,3}. In the case of *In*, both sexes have poor viability and females also suffer a decline in fertility. So, as in the case of the sickle-cell gene, homozygosity for *In* is something to be avoided if possible. However, unlike the case of the sickle-cell gene, homozygosity of *In* is frequently avoided. It is this that makes *In* so very remarkable.

As noted above, if a female heterozygous for the driver (*In*/+) mates with a male homozygous for the wild type (+/+) an excess of progeny inherit the driving gene from their mother (85 per cent of the two-month-old progeny are *In*/+). But if the male is homozygous for the driver (*In*/*In*), then the opposite occurs, most progeny inheriting the wild-type gene from their mother (65 per cent of the two-month-old progeny are *In*/+) (see figure). This is extraordinary, as almost all progeny in the latter cross ought to be drive homozygotes (*In*/*In*) given the transmission bias in favour of *In*.

An obvious interpretation of the low frequency of homozygotes in this cross is that the drive homozygotes are produced but die early in life and are not seen. Indeed, many homozygotes of both sexes die during the first two months of life². However, by counting the number of *corpora lutea* and implantation sites of the females, Agulnik and colleagues could estimate how many embryos were produced but died before they could be counted¹. Against expectations, they discovered that mortality rates were no different to those of normal crosses.

Importantly, the rate was far too small to account for the high proportion of *In*/+ offspring.

Agulnik *et al.* propose that these patterns can be explained if drive in female heterozygotes is dependent upon the genotype of the fertilizing sperm. In heterozygous females drive is normal if wild-type sperm fertilize the egg, producing an excess of heterozygotes. If the fertilizing sperm contains the *In* chromosome, however, then drive is abolished and segregation follows a near-normal mendelian pattern (in fact there might even be a bias in favour of the wild-type chromatid). Further support for the notion that sperm genotype affects oogenesis comes from the finding that such a model correctly predicts the frequency of each progeny genotype in crosses in which both parents are heterozygous¹.

If, after fusing with the oocyte, sperm are affecting oogenesis, it must be that the second meiotic division in the oocyte does not occur until after sperm entry. In mice,

as in many other species⁶, that division is suspended until the sperm penetrates the egg. The problem then remains how the sperm communicates its genetic composition to the chromatids undergoing meiosis. Agulnik and colleagues discuss two possibilities¹. First, sperm structures might directly affect the process of meiosis, perhaps by modifying the meiotic spindle; second, the sperm might indirectly affect proceedings by emitting some signal that elicits a response in the oocyte. Agulnik *et al.* argue that, as the oocyte meiotic structures have already been formed by the time the sperm penetrates, it is more reasonable to suppose that the *In*-bearing sperm are emitting some signal. This hypothesis has yet to be tested.

Although the mechanism is unclear, the new work shows that meiotic segregation, and hence zygote fitness, can be adjusted in light of the sperm's genetic constitution. Is this however an instance of adjustment of meiosis to prevent zygotic failure? Can we be sure that the adjustment is made to avoid the production of *In* homozygotes because such homozygotes are less fit? One could argue that the effect is simply a curious side-effect of a curious gene, rather than an adaptation to minimize the gene's deleterious effects. To establish the latter will require analysis of this system in the wild, so as to show that the effect is beneficial for the gene controlling the adjustment. But for the time being it seems reasonable to suppose that the effect is adaptive, at least for genes unlinked to *In*.

How common might the *In* effect be? At first sight it seems it must be rare: were it not so it would surely have been seen before. But many cases of non-mendelian birth ratios are put down to differential embryonic mortality even though there is no firm evidence of early mortality, including mortality of unrecoverable embryos. Maybe choice of the maternal chromatid most compatible with the sperm pro-nucleus is more common than previously thought. Whether or not this wild conjecture has any substance, a few pages of genetics textbooks must now be re-written and we may also have to reconsider just how *In*-telligent genes can be. □

Andrew Pomiankowski is in the Department of Genetics and Biometry, University College London, London NW1 2HE, UK. Laurence D. Hurst is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Black holes and information

William A. Hiscock

THEORETICAL physicists are fortune tellers who welcome the experimental testing of their predictions; but to predict the future, the current physical state of the system must be known with precision. When matter collapses to form a black hole, it seems that an enormous amount of information is lost to the Universe, forever hidden behind the impenetrable veil of the event horizon. Worse, as the black hole evaporates because of quantum particle creation¹, it seems that the information trapped in the black hole may disappear from our Universe altogether. If the information truly vanishes, and cannot be recovered, then a crisis looms for our present understanding of quantum physics: determinism (within the context of quantum mechanics) is lost. Recent work by Callan, Giddings, Harvey and Strominger² examines a possible escape from this plight, and has stimulated much interest in this apparent clash of quantum physics and gravitation.

Description

To describe a star completely, a huge amount of information must be specified: total mass and angular momentum, chemical composition, temperatures, pressures, gravitational multipole moments and electromagnetic fields, and so on down to the actual quantum states of the roughly 10^{57} nuclei contained in the star. After the star collapses to become a black hole, the various black-hole uniqueness theorems³ guarantee that the hole can be completely described by at most four parameters — the mass, angular momentum, electric charge, and (if magnetic monopoles exist) magnetic charge. All the other information that was originally contained in the star is inaccessibly hidden inside the event horizon.

Shortly after his original discovery of quantum pair production around black holes, Hawking pointed out⁴ that the perfectly thermal, black-body spectrum of the emitted particles precludes them from carrying off any of the information trapped in the black hole. When the black hole ultimately vanishes as it loses the last of its mass, all the information that was trapped inside (presumably in the singularity) apparently vanishes entirely from our Universe.

Despite the statistical nature of predictions in quantum theory, the underlying wavefunction that describes the quantum state of a system is believed to evolve in such a way that information is neither created nor destroyed. If the picture developed by Hawking is correct, then gravitational collapse violates one of the most fundamental tenets of quantum theory.

Theoretical physicists studying this predicament have identified three basic possibilities for the final demise of a black hole. In the first, the information trapped in the black hole is actually carried away in the radiation created, in subtle correlations that would not have been evident in Hawking's original calculations. In the second case, the black hole does not evaporate completely, but leaves behind a 'remnant' of roughly the Planck mass, 10^{-5} grams, which somehow contains all the original information. In the third, the black hole evaporates completely and information is irretrievably lost.

In the 15 years following Hawking's discovery of black-hole evaporation, little progress was made on determining which of the above pictures (if any) is correct. The primary difficulty is that the last moments of black-hole evaporation, when the black hole's mass is comparable to the Planck mass, must be described by a quantum theory of gravity. Despite decades of effort, and the intriguing mathematical properties of superstring theories, it cannot be said that we have an adequate theory of quantum gravity today. This difficulty, of course, also gives the problem much of its allure: perhaps by carefully studying the endpoint of black-hole evaporation, we can start to see how gravitation and the quantum principle will ultimately be reconciled.

During the past year, there has been a flood of new work on the question of information loss in black-hole evaporation. The stimulus was the 1992 paper of Callan *et al.*², which pointed the way to a simplified model of black-hole evaporation in which the question could be studied with some precision. There were two key points that led to their simplified model.

Dimensions

First, Callan *et al.* noted that for spherically symmetric black holes the essential physics of evaporation is two-dimensional (radius and time). The angular degrees of freedom are not expected to contribute to the physics in any qualitatively important fashion, although precise quantitative results would certainly be different in two- and four-dimensional models. This has long been known, and there is a long history of studies of black-hole evaporation in two dimensions⁵. But earlier studies were thwarted by the fact that Einstein's equations of general relativity are vacuous in two dimensions; the Einstein tensor, which describes how matter curves the spacetime, is always zero in two dimensions. Callan *et al.* avoid this problem by working in a slightly different theory of 'dilaton' gravity, which is not

1. Agulnik, S. I., Sabantsev, I. D. & Ruvinsky, A. O. *Genet. Res.* **61**, 97–100 (1993).

2. Agulnik, S. I., Agulnik, A. I. & Ruvinsky, A. O. *Genet. Res.* **55**, 97–100 (1990).

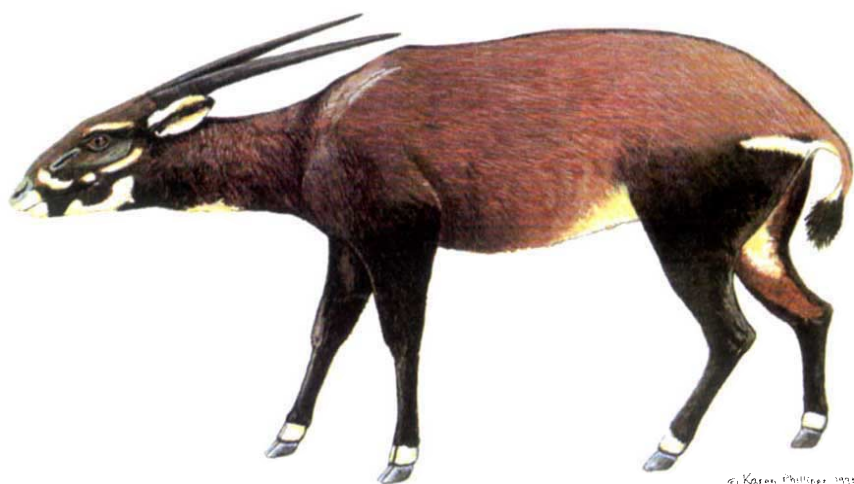
3. Lyttle, T. W. A. *Rev. Genet.* **25**, 511–557 (1991).

4. Agulnik, S. I., Sabantsev, I. D., Orlova, G. V. & Ruvinsky, A. O. *Genet. Res.* **61**, 91–96 (1993).

5. Agulnik, S. I., Borodin, P. M., Gorlov, I. P., Ladygina, T. Y. & Pak, S. D. *Heredity* **65**, 265–267 (1990).

6. John, B. *Meiosis* (Cambridge University Press, 1990).

Reconstructed from skin and bone



JUST when one might think that people have looked behind every tree on the planet for hitherto undescribed mammals, one emerges from the forests of Vietnam. In a classic piece of old-fashioned zoology, V. V. Dung *et al.* (page 443 of this issue) describe a new species of bovid solely on the basis of local hunting trophies — skulls, teeth and skins. The picture above is an artist's reconstruction, and science has yet to scrutinize a living specimen of *Pseudoryx nghetinhensis*. A few hundred animals may exist in the montane forests of the Vietnamese–Laotian border, and its status must be considered as endangered.

The case of another rare even-toed ungulate, the Chacoan peccary (*Catagonus wagneri*), may be a guide to its fate. This pig-like animal had been known exclusively from fossils, until a living population was discovered in the Chaco region on the borders of Bolivia, Paraguay and Argentina in the 1970s. Until the arrival of *Pseudoryx*, the peccary was the latest and last large mammal to have been described — but hunting and habitat destruction could wipe it out before this century closes.

Others are even less lucky: the gazelle *Gazella ruffina* is known only from skins garnered from nineteenth-century Algerian souks. Undeterred by the subsequent non-appearance of *G. ruffina*, Dung and colleagues intend to hunt for living specimens of *Pseudoryx*. The animal gets its name from the oryx, which has similar long, graceful horns. But wild populations of the Arabian oryx, *Oryx leucoryx*, were wiped out in the 1970s, and the species was saved only by captive breeding.

The benign hunt for *Pseudoryx* should have a similar aim — to rescue the animal and transfer it to a safe haven. Indeed, the animal could have chosen better places to emerge from hiding than Indochina. This is the haunt of another shy bovid, the kouprey (*Bos sauveli*), described in 1937. It numbered about 1,000 animals in 1940, but was down to 100 in 1969. "Since its discovery," writes Ronald Nowak (in *Walker's Mammals of the World*, 5th edn, Johns Hopkins University Press, 1991), "its limited range has been the scene of almost constant human warfare." The kouprey was reported to be hanging on, albeit grimly, in 1986. Reports of its demise will be neither an exaggeration nor, sadly, a surprise.

H.G.

precisely general relativity, although it is closely related to both general relativity and the low-energy limit of superstring theories. This theory, unlike general relativity, has dynamics in two dimensions.

The second stratagem used by Callan *et al.* is to study extreme charged black holes, that is, black holes whose charge is equal in magnitude to their mass (in 'natural units' where Planck's constant, Newton's constant and the speed of light are all set equal to unity). Such black holes have zero temperature, and hence are stable against evaporative particle creation, ignoring charged particle creation. If such a black hole absorbs an infalling

neutral particle, the hole's mass will be increased while the charge remains fixed. This lowers the charge-to-mass ratio and raises the temperature of the black hole above zero. The black hole then emits one or several particles, and thereby loses enough mass to return to its stable extremal ground state. Viewed from afar, this appears to be a scattering process: there is a well defined incoming state (the infalling particle) and a well defined outgoing state (the created particles). The most powerful tools of quantum field theory, which are explicitly designed to deal with perturbative scattering, may then be brought to the study of this situation.

Within the context of this model, Callan *et al.* found that black holes did not form at all; the collapsing matter radiated away energy at a sufficient rate to prevent the formation of an event horizon. If an event horizon never forms, then the evolution of the quantum mechanical wavefunction from the past to the future remains deterministic.

On realizing that tractable two-dimensional models of black-hole evaporation exist, quantum theorists have rapidly pressed them into service. Of the immense amount of new work, most has focused on understanding the possible physics of black-hole remnants, the massive (10^{-5} gram) objects, 10^{-33} cm in size, which may contain all the information in the original collapsing star. If such remnants exist, they are surely the most compact information storage containers in nature. A model of such a remnant consisting of a tiny extreme charged black hole has been termed a 'cornucopion' because of the need to have a tremendous amount of information stored in an incredibly small object⁶.

What then is the present state of our knowledge concerning the endpoint of black-hole evaporation? As the two-dimensional studies progress, the possibility that the information in the collapsing matter escapes through subtle correlations in the Hawking radiation seems increasingly unlikely. Further insights into the behaviour of remnants are emerging; however, there are still serious difficulties concerning possible copious pair creation of these objects as fundamental particles. Finally, Hawking's original notion that the information is in fact destroyed by black-hole evaporation remains a real possibility, although one that will not be palatable to many theoretical physicists until other possible resolutions are completely ruled out.

Since its discovery, black-hole evaporation has been the process whose theoretical study has provided the greatest insights into the physical nature of quantum gravity. The present enthusiasm for this field results from the realization that out of the apparent crisis in theoretical physics posed by information loss in black holes there may arise a new understanding of physics at the most fundamental levels. □

William A. Hiscock is in the Department of Physics, Montana State University, Bozeman, Montana 59717, USA.

1. Hawking, S. W. *Nature* **248**, 30 (1974); *Commun. math. Phys.* **43**, 199–220 (1975).
2. Callan, C. G., Giddings, S. B., Harvey, J. A. & Strominger, A. *Phys. Rev. D* **45**, R1005–R1009 (1992).
3. Wald, R. *General Relativity*, 322–324 (University of Chicago Press, 1984).
4. Hawking, S. W. *Phys. Rev. D* **14**, 2460–2473 (1976).
5. Davies, P. C. W., Fulling, S. A. & Unruh, W. G. *Phys. Rev. D* **13**, 2720–2723 (1976).
6. Banks, T., Dabholkar, A., Douglas, M. R. & O'Loughlin, M. *Phys. Rev. D* **45**, 3607–3616 (1992).

Heavy carbon dioxide

Ralph F. Keeling

THE isotopes of oxygen in atmospheric carbon dioxide have long suffered from neglect — as one veteran of the business has put it, “figuring out the carbon isotopes was enough work without having to bother about the oxygen isotopes as well”. Well, people are finally starting to bother. And as Farquhar *et al.* now show (on page 439 of this issue¹), the oxygen isotopes should provide valuable information on exchanges of carbon between the Earth’s atmosphere, oceans and terrestrial biota.

Globally, there is much more oxygen in water and oxygen gas than in CO₂. Oxygen atoms in CO₂ are not exchanged with water vapour or oxygen gas; rather, such exchange takes place only when CO₂ is dissolved in liquid water and only when the dissolved CO₂ is hydrolysed as carbonic acid. The hydrolysis reaction is so slow that a CO₂ molecule must remain dissolved in water for at least 20 seconds to have much chance of exchanging oxygen atoms with the water. In general, allowing for kinetic and equilibrium fractionation, the ratio of oxygen isotopes (¹⁸O and ¹⁶O)

in atmospheric CO₂ should follow the ¹⁸O/¹⁶O ratio of the liquid water reservoirs with which the atmosphere makes most frequent and prolonged contact.

Those ratios vary considerably. Cloud water, rain water and soil water are depleted in H₂¹⁸O relative to sea water because fractional distillation in the atmosphere produces lower ¹⁸O/¹⁶O in precipitation at higher latitudes. Leaf water is enriched relative to soil water by evapotranspiration because H₂¹⁸O evaporates less readily than H₂¹⁶O.

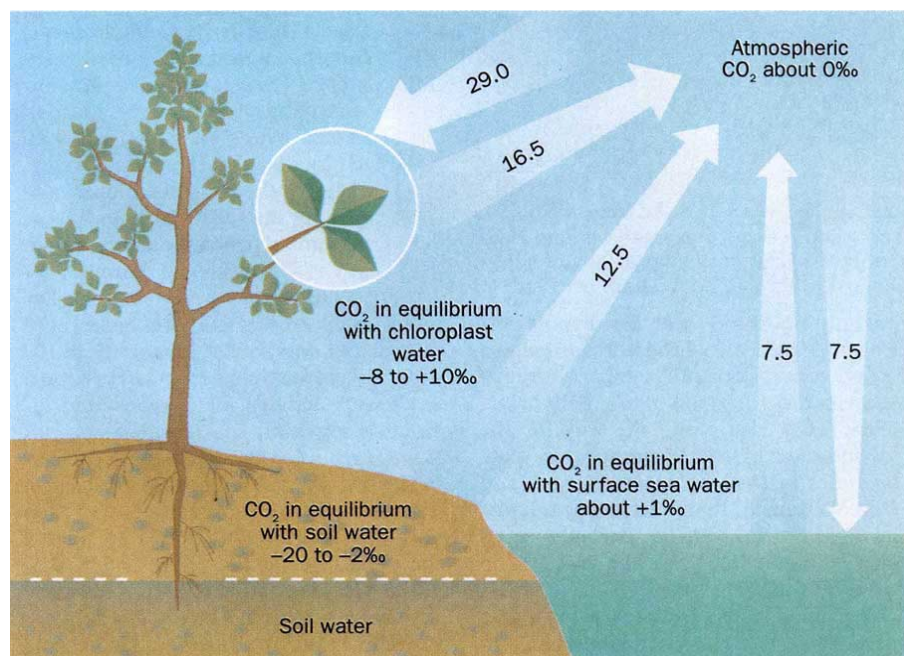
Previously, the oceans were believed to be the primary reservoir for exchange of oxygen atoms in CO₂ because early measurements showed that the ¹⁸O/¹⁶O ratio of atmospheric CO₂ was very close to the value expected for equilibrium with surface sea water^{2,3}. Subsequent measurements, however, revealed large-scale patterns in ¹⁸O/¹⁶O of CO₂ that were out of sync with what would be expected from calculations based on the temperature variations at the sea surface⁴⁻⁶. Particularly interesting was the observation by

Francey and Tans⁶ that the ratio of ¹⁸O to ¹⁶O in CO₂ is much lower at high latitudes in the Northern Hemisphere. Maintenance of this deficit in spite of vigorous atmospheric mixing means that the turnover time of oxygen atoms in CO₂ must be as short as two years. This is much less than the exchange time for CO₂ with the oceans (about 10 years) and, in fact, is shorter than the exchange time for carbon atoms in CO₂ with any reservoir.

What process could exchange the oxygen atoms but not the carbon atoms in atmospheric CO₂? Francey and Tans speculated that a large fraction of the CO₂ that enters leaf cells was leaking back out to the atmosphere after exchanging oxygen atoms with leaf water. Although the typical residence time for a CO₂ molecule in leaf water is less than one second, oxygen atom exchange could occur with the help of the enzyme carbonic anhydrase, which is present in the chloroplasts of leaf cells and dramatically speeds up the hydrolysis reaction.

Farquhar *et al.* now show that water in the chloroplasts, where oxygen exchange occurs, has effectively the same isotopic composition as water at the evaporating sites in the leaves. This allows them to quantify the ideas of Francey and Tans with the use of a relatively simple model relating the isotope exchanges to the gross primary production of plants. Even so, the model necessarily requires keeping track of a large number of variables — the isotopic composition of water in the soil and air, leaf temperature, water vapour pressure in the air, the fraction of CO₂ that diffuses back out of the leaf, and gross primary production — and producing maps of these variables over the surface of the Earth. Farquhar *et al.* test their model by showing that it produces the correct ¹⁸O/¹⁶O ratio for bulk atmospheric CO₂ (see figure), and then use the model to show qualitatively how the deficit in ¹⁸O/¹⁶O of CO₂ at northern high latitudes arises from the decrease in ¹⁸O/¹⁶O of soil and leaf water at these latitudes.

The logical next step for the authors is to combine their model with numerical simulations of large-scale air motions so they can test the predictions against observed variations in ¹⁸O/¹⁶O of CO₂. One area needing more work is the iso-



The global cycle for oxygen isotopes of atmospheric CO₂. Some 55% of the CO₂ that diffuses into the chloroplasts of leaf cells diffuses back out to the atmosphere after exchanging oxygen atoms with the water in the chloroplasts. This exchange drives the ¹⁸O/¹⁶O ratio of atmospheric CO₂ towards the value expected from equilibrium with chloroplast water. The remaining 45% of CO₂ is taken up by photosynthesis, and eventually returns to the atmosphere through respiration or decomposition in the soil. This fraction, known as gross primary production, drives the ¹⁸O/¹⁶O ratio of atmospheric CO₂ towards equilibrium with soil water. (Note that the kinetic isotope fractionations caused by diffusion into leaves and out through soil tend to cancel.) Roughly 45% of the oxygen atoms in atmospheric CO₂ comes from chloroplast water at an average isotopic composition of +5‰, 34% come from soil water at an average of -7‰, and 21% come from sea water at +1‰. This combination yields atmospheric CO₂ at 0‰. The annual CO₂ fluxes shown are in units of 10¹⁵ moles of carbon. The ¹⁸O/¹⁶O ratios of CO₂ are expressed as deviations ($\delta = (^{18}\text{O}/^{16}\text{O})/(^{18}\text{O}/^{16}\text{O})_{\text{PDB}} - 1$) from the ratio of the carbonate standard Pee-Dee Belemnite (PDB). The deviations are multiplied by 1,000 and expressed in per mil (‰).

1. Farquhar, G. D. *et al.* *Nature* **363**, 439–443 (1993).
2. Keeling, C. D. *Geochim. cosmochim. Acta* **24**, 277–298 (1961).
3. Bottinga, Y. & Craig, H. *Earth planet. Sci. Lett.* **5**, 285–295 (1969).
4. Mook, W. G., Koopmans, M., Carter, A. F. & Keeling, C. D. *J. geophys. Res.* **88**, 10915–10933 (1983).
5. Friedli, H., Siegenthaler, U., Rauber, D. & Oeschger, H. *Tellus* **B39**, 80–88 (1987).
6. Francey, R. J. & Tans, P. P. *Nature* **327**, 495–497 (1987).
7. Yakir, D., Berry, J. A., Giles, L., Osmond, C. B. & Thomas, R. in *Scaling Physiological Processes: Leaf to Globe* (eds Ehleringer, J. R. & Field, C. B.) 323–338 (Academic, San Diego, 1993).
8. Yakir, D., Berry, J. A., Giles, L. & Osmond, C. B. in *Stable Isotopes and Plant Carbon Water Relations* (eds Ehleringer, J. R., Hall, A. E. & Farquhar, G. D.) (Academic, San Diego, in the press).

topic composition of chloroplast water because some studies have indicated that chloroplast water is not always as enriched in ^{18}O as the water at the evaporating sites in leaves^{7,8}. Eventually, the aim is to use $^{18}\text{O}/^{16}\text{O}$ data to tie down uncertainties in the exchange of CO_2 and water with terrestrial ecosystems.

One such application proposed by Farquhar *et al.* is to use the data to constrain estimates of net uptake or release of carbon from terrestrial ecosystems.

This is likely to prove difficult because such net fluxes produce only small variations in $^{18}\text{O}/^{16}\text{O}$ which will probably be masked by larger variations caused by other processes. The unique value of the oxygen isotopes is their sensitivity to rates of gross primary production and rates of diffusion through stomata, as these exchanges can produce large variations in the isotopes without producing detectable changes in the carbon isotopes or the total concentration of CO_2 (refs 6, 7).

Perhaps the most remarkable outcome of this work is the confirmation that the minuscule amount of water in chloroplasts actually dominates the isotopic signature of oxygen atoms in atmospheric CO_2 . Score another point for those who argue that biology rules the chemistry of the Earth. □

Ralph F. Keeling is at the Scripps Institution of Oceanography, La Jolla, California 92093, USA.

OBITUARY

John Tuzo Wilson (1908–1993)

THE 1960s saw a revolution in the Earth sciences: the emergence of the concepts of seafloor spreading and plate tectonics and the final ascendancy of a mobilistic view of the solid Earth, originally expressed in terms of continental drift. Few people were at the forefront of this revolution throughout the decade. Arguably only one was — J. Tuzo Wilson, the Canadian geophysicist, who died in Toronto on 15 April.

In 1960 Wilson was himself a recent convert to the concept of continental drift and immediately set about preaching the word with all the zeal of a new convert. Belief in such heresy was then virtually unknown in North America. Unlike most of his colleagues, Wilson was receptive to the idea of Harry Hess, first circulated as a preprint in 1959, that the ocean floors are relatively young in geological terms and are constantly being created at the mid-ocean ridges and destroyed at the marginal ocean trenches, a concept immediately extended, and dubbed seafloor spreading, by Bob Dietz.

Undoubtedly Wilson's most significant contribution to these evolving ideas was his recognition, in 1965, that seafloor spreading and continental drift imply that there must be a third major structural element in addition to the ridges and trenches. This he termed a transform fault. In places, ridges or trenches end abruptly and 'transform' into major strike-slip faults. In contrast to the ridges and trenches, transform faults are conservative features implying neither creation nor destruction of crust. They are also active only between the ridge or trench terminations they join, thus providing an elegant model for terminating major strike-slip faults.

In common with the suggestion by Drummond Matthews and myself that seafloor spreading might act as a tape recorder, recording reversals of the Earth's magnetic field, the concept of transform faults had the virtue of being testable by further work. Thus in 1966,

Lynn Sykes presented improved epicentral locations and focal-mechanism solutions for earthquakes occurring on certain proposed transform faults, which convincingly confirmed the hypothesis. This came close on the heels of new magnetic data lending remarkable sup-

port to the development of passive and active margins, and mountain belts, a sequence now commonly termed the Wilson cycle.

Wilson's road to Damascus conversion and his most fruitful contributions to his subject came quite late in his career. In this regard he did not have a very helpful start. In the early 1930s he held a Massey Fellowship at Cambridge, where he came under the influence of Harold Jeffreys, and later in the decade he joined the Canadian Geological Survey and was put to work on the rather stable Canadian shield. Between the two he did a PhD at Princeton where Hess was a young lecturer; but this was nearly thirty years before Hess formulated seafloor spreading. During the Second World War Wilson served with distinction in the Canadian Army, rising to the rank of Colonel, and gaining the OBE. In 1946 he was appointed professor of geophysics at the University of Toronto.

Tuzo was a tireless traveller and speaker. It was said that he lectured at over 200 universities in more than 100 countries, something he seemed to regard as a leisure activity, along with sailing his Hong Kong junk in Georgian Bay off Lake Huron. He also had a great interest in the history and philosophy of science, and in the public understanding of science. Thus in the late 1960s he related the revolution in the Earth sciences to Kuhn's ideas on the structure of scientific revolutions; and in 1974, on retiring from the University of Toronto, he became Director of the Ontario Science Centre, the first science exploratory.

Tuzo Wilson was a big man; in stature, in his magnanimity, on the Canadian science scene, and internationally in the Earth sciences. His seminal contributions to our new view of the Earth will be long remembered.

Fred Vine

Fred Vine is in the School of Environmental Sciences at the University of East Anglia, Norwich, NR4 7TJ, UK.



Tuzo Wilson (centre) being presented with the AMSOC Albartross award by Sir Edward Bullard at a conference held in Wilson's honour in Toronto, May 1979. Left, Art Maxwell.

port to Matthews and my hypothesis. Together these two developments led to a mass conversion of North American geoscientists to spreading and drift in late 1966. They also paved the way for the formulation of the kinematics of plate tectonics in 1967 by Jason Morgan and Dan McKenzie. In this, transform faults have a crucial role in defining small circles about the eulerian pole of a rotation for plate pairs.

Wilson was also in the vanguard of a number of other key developments. In 1963 he suggested that the Hawaiian and other island chains might have formed by movement of the crust over an essentially stationary 'hotspot' or volcanic centre. This idea was subsequently developed by Morgan into the hotspot or plume model for 'absolute' plate motions, whereby the motion of plates with respect to the deep interior of the Earth can be determined for the past 200 million years. In the late 1960s Wilson developed the concept of repeated opening and closing of ocean basins throughout geological time with its

A rational attack on influenza

Garry Taylor

THE influenza virus has been successfully evading the immune system for centuries, making vaccines against it modestly effective at best. Any drug therapy must target parts of the virus that remain invariant among strains, one such part being the active site of the influenza virus neuraminidase. On page 418 of this issue¹, von Itzstein and colleagues describe potent inhibitors of this enzyme which have been rationally designed and which are effective in inhibiting flu virus replication in animal models. These results will excite and encourage those involved in rational drug design, in which knowledge of the chemistry and structure of a receptor molecule (an enzyme, for example) is used to design ligands with optimized chemical interactions and binding affinity. But it has to be said that several more steps have to be taken before the promise of this particular development is realized.

Do we need to prevent flu? Each year as winter approaches, millions of shots of vaccine are given mainly to people who are especially susceptible to the virus (the elderly, asthmatics, diabetics). The vaccine is a cocktail of recent influenza A and B strains and affords some protection against any epidemics caused by antigenic drift of these vaccine strains. Nevertheless, in recent history thousands have died from influenza infection; in the 1957–58 epidemic, for instance, 1 in 300 of those over 65 died.

The problem is that, on occasion, the virus undergoes a remarkable change in its antigenic appearance by genetic reassortment of two very different strains infecting one animal or bird; in particular, ducks provide a vast gene reservoir of influenza virus which could become part of a reassortant virus. A pandemic may follow, often intriguingly originating in China, with devastating effects: 1890, 1900, 1918 (Spanish flu), 1957 (Asian flu) and 1968 (Hong Kong flu) were years characterized by widespread mortality. The 1918–19 pandemic killed 20 to 40 million people worldwide; it could happen again. Influenza has also had devastating effects on fowl, horses, swine and seals.

Chemotherapy or chemoprophylaxis of influenza has to date been limited to derivatives of amantadine, which blocks the ion channel of the viral M2 protein, but side effects have limited their usefulness. The determination, just over ten years ago, of the atomic X-ray structures of the two viral surface glycoproteins,

haemagglutinin² and neuraminidase³ (sialidase), started the quest for rationally designed inhibitors of these proteins. Haemagglutinin recognizes sialic (neuraminic) acid as the receptor for cell attachment; sialidase removes sialic acids from glycoconjugates, and is thought to aid virus penetration of the mucosal lining of the upper respiratory tract and later to



Flu kills. The Austrian artist Egon Schiele painted this picture of *Die Familie* in 1918, the man being a self-portrait. Months later Schiele's own immediate family had become victims of flu. He and his pregnant wife died within days of each other after contracting the Spanish flu that swept the world at the end of the First World War.

process progeny virus particles to stop virus agglutination.

In 1969, a synthetic derivative of sialic acid called Neu5Ac2en, or DANA, was shown to inhibit bacterial, viral and mammalian sialidases⁴ *in vitro* with similar inhibition constants ($K_i \sim 10^{-5}$ M). Many analogues of DANA have since been synthesized^{5–7}, but none have K_i values less than micromolar, and none show any significant activity *in vivo*. Following the determination of the influenza neuraminidase structure in 1982 by Colman, Laver and Varghese, a small biotechnology company, Biota, was formed in Melbourne to exploit the structural results.

The latest fruits of this work are now presented by von Itzstein *et al.* They have probed the active site in detail to predict sialic-acid derivatives that have substantially improved binding affinities. The crystallographic analysis of the binding of sialic acid or DANA revealed an empty pocket near the O4 hydroxyl on the sugar

ring. From the neighbouring acidic amino acids, it seemed that substitution of O4 with an amino or guanidino group should fill the pocket. The synthesized compounds bound as predicted, with K_i values in the nanomolar range (two to four orders of magnitude better than DANA), and in accord with the formation of an additional salt bridge. This pocket appears to be an invariant feature of existing influenza A and B virus neuraminidases, hence the effectiveness of the compounds *in vitro* against two A strains and one B strain. Furthermore, the com-

pounds are pretty specific for influenza virus sialidases, having only millimolar inhibition constants with sialidases from *Vibrio cholerae*, sheep liver, parainfluenza and human lysosomes. The studies with mice and ferrets show good inhibition of virus replication, when the new compounds are administered as a nasal spray, and that they are far more effective than amantadine.

So do we now have an effective way to prevent flu? 'Not yet' is the answer. Most obviously, the efficacy of the compounds is yet to be demonstrated in humans. Also the occurrence of drug-resistant mutants, when the highly adaptable influenza virus comes under the pressure of a new evolutionary force, is a possibility. One puzzle is why the empty pocket in the influenza neuraminidase active site has been preserved. von Itzstein *et al.* predict that the non-influenza sialidases probably do not possess this pocket, and this is being confirmed by structural studies on bacterial sialidases⁸. Interestingly, one of these sialidases from *Salmonella typhimurium* catalyses identical reactions and even shares, with the influenza enzyme, a kinetic preference for certain types of glycosidic linkages⁹. Perhaps then, the flu virus may become resistant to these new compounds.

Sialidases are ubiquitous¹⁰. They have long been implicated in the pathogenesis of many bacteria, for example those

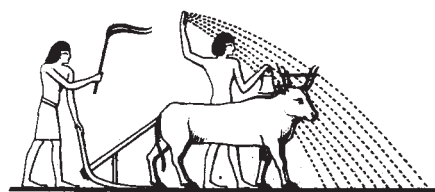
1. von Itzstein, M. *et al.* *Nature* **363**, 418–423 (1993).
2. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
3. Varghese, J. N., Laver, W. G. & Colman, P. M. *Nature* **303**, 35–40 (1983).
4. Meindl, P. & Tuppy, H. *Hoppe-Seyler's Z. physiol. Chem.* **350**, 1088–1092 (1969).
5. Meindl, P., Bodo, G., Palese, P., Schulman, J. & Tuppy, H. *Virology* **58**, 457–463 (1974).
6. Flashner, M., Kessler, J. & Tanenbaum, S. W. *Archs Biochem. Biophys.* **221**, 188–196 (1983).
7. Glänzer, B. I., Györgydeák, Z., Bernet, B. & Vasella, A. *Helv. chim. Acta* **74**, 343–369 (1991).
8. Taylor, G., Vimr, E., Garman, E. & Laver, W. G. *J. molec. Biol.* **226**, 1287–1290 (1992).
9. Hoyer, L. L., Roggentin, P., Schauer, R. & Vimr, E. R. *J. Biochem.* **110**, 462–467 (1991).
10. Corfield, A. P., Lambre, C. R., Michalski, J.-C. & Schauer, R. in *Conferences Philippe Laudat 1991* 113–134 (Institut National de la Santé et de la Recherche Médicale, INSERM, Paris, 1992).

involved in cholera, meningitis and septicaemia. The paramyxoviruses, including those responsible for mumps, measles, Newcastle disease and parainfluenza, have a single surface glycoprotein with both haemagglutinin and sialidase functions. Several sialidases have been found in man, and they carry out important functions related to cell-cell recognition through sialoglycoconjugate metabolism. The specificity of the new anti-influenza compounds for one type of sialidase gives hope that disease specific inhibitors of this enzyme may be developed. On the other hand, we know very little about mammalian sialidases, and designers of any sialidase-based inhibitors would have to avoid compromising the endo-

genous activities of these enzymes.

Influenza derived its name from fourteenth-century Florence, when an unusual conjunction of planets was thought to 'influence' epidemics of coughs, colds and fevers. Our knowledge of the influenza virus, and of our immunity to it and its epidemiology, have come a long way since then, but conjunction of planets may still be the best method we have of predicting when a new pandemic strain might emerge. With the work of von Itzstein *et al.*, we have gone some way towards being prepared for that day. □

Garry Taylor is in the Department of Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.



alluvium. The oldest Predynastic remains that can be generally reached in the delta belong to what is known as the Buto-Maadi Culture. This dates to roughly the middle of the fourth millennium BC.

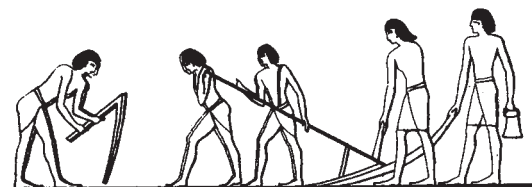
Stanley and Warne's results have significant bearing on our understanding of the origins of food production in Egypt as a whole. The delta seems the most likely place for domestic crops and animals to have been introduced into Egypt from the Near East (where the first cereals and animals had been domesticated by the end of the tenth millennium BP, before present), so that one might have expected Egypt to have remained a strictly hunter-gatherer culture until the delta became suitable for agriculture. Instead we find indications that the people of the southern Western Desert of Egypt were initiating their own 'Neolithic revolution' as early as the mid-tenth millennium BP. Cattle remains and pottery have been found at sites of this date in the Bir Kiseiba area⁴. Today the region can go without rain for 20 years, but even during the comparatively moist earlier Holocene (the postglacial period

ARCHAEOLOGY

Rise of the Nile delta

D. L. Holmes

A PICTURE of human civilization carried along on a tide of global climate change is painted in a report by Stanley and Warne, on the emergence of agriculture in the Nile delta (page 435 of this issue¹). The authors find that the deposition of cultivatable silts of the delta began at around 6500–5500 BC, after sea level had risen over 100 m above its glacial minimum.



The establishment of agriculture was the first vital step in the development of Egyptian civilization. It was long believed that the Nile delta in late prehistoric times was a vast uninhabitable swamp. No early farming settlements were known (except for Merimda at the delta's western edge) and none was expected. Subsequently it was realized that the area had not been so inhospitable, although it was argued that any prehistoric sites would have been deeply buried under thousands of years worth of Nile alluvium.

But this was another idea to bite the dust. Many Predynastic delta sites are now known and several have been excavated since the late 1970s. The geological investigations of Stanley and Warne provide an indication of when the delta would have first been available to agricultural settlers. From an extensive series of radiocarbon-dated drill-cores, they show that the present Nile delta with its rich silts and muds

started to form only after 6500–5500 BC. This they link to a deceleration in sea-level rise, rather than to the effect of regional climate on Nile flood levels. From the Last Glacial Maximum (about 18,000 years ago) until 8,000 years ago, sea level rose rapidly from a global minimum of approximately 125 m below the present sea level to 16 m below the present level. Only when the rise in sea level decelerated and the river gradient became sufficiently reduced did the silt-rich deposits begin to accumulate, and the delta area become suitable for farming.

The estimate that deposition of flood-plain silts started around 6500–5500 BC fits well into the archaeology of the delta. Merimda, the delta's oldest known Neolithic site (the term Predynastic is normally reserved for the period following 4000 BC, when copper was introduced) was probably first occupied around 4900–5000 BC². If at least a few centuries are allowed for the Nile silts to build up, there is little prospect of finding evidence of much older farming communities in the Nile delta.

Merimda, however, remains the only known Neolithic delta site. There are indications of an occupation of comparable age at Minshat Abu Omar in the eastern delta³, where hand-auguring resulted in the recovery of several potsherds from a level immediately below an organic-rich mud which was radiocarbon-dated to 4700 BC. But this locality is unlikely to be excavated as it lies 6 m or more below the present water table. Any other delta sites of Neolithic vintage are also likely to be buried well below the water table. When Hermann Junker discovered Merimda in the 1920s, he was lucky: the site lies on the desert margin of the delta about 3 m above the delta



beginning at 10,000 BP), annual rainfall was probably no more than around 50 mm (ref. 5) — not nearly enough to sustain wild cattle, so that the cattle were arguably domestic.

Later on, at around 8000 BP, there is evidence suggesting that sorghum was being cultivated⁶. But this was a crop that was to be adopted in the Sudan and sub-Saharan Africa rather than in the prehistoric Egyptian Nile valley. Then came domestic sheep and goats from the Near East. The oldest North African representatives, from the mid-eighth millennium BP, were in the Central Sahara⁷. In contrast, the early Holocene occupants of the Nile valley and the Fayoum depression (70 km southeast of Cairo) seem to have depended on wild resources.

The suggestion is that people from the desert migrated to the Nile valley as conditions became increasingly arid during the middle Holocene^{8,9}: there are abundant sites in the Western Desert dating from before about 5000 BC, but a sharp decline afterwards, whereas Nile valley sites become much more frequent

Insets: ploughing, sowing and reaping, after Egyptian wall paintings (Mary Evans).

RÉSUMÉ

Negative response

THE notion that raising a voltage above a threshold can cut the resulting current, so-called negative differential resistance, is now being applied in modern electronics. But the notion of negative absolute conductivity, so that currents run against a voltage, is rather more bizarre. First discussed in the 1960s, it failed to appear in experiments because thermal effects of bulk materials erased the quantum mechanisms that should generate the effect. But modern techniques that allow the construction of ultrapure, one-dimensional semiconductor channels should change all that, say R. Mickevičius *et al.* (*Appl. Phys. Lett.* **62**, 1970–1972; 1993), who have run Monte Carlo simulations to substantiate their claim. The effect should arise in illuminated photoconductors owing to asymmetric interactions between electrons and lattice vibrations.

Fowl flu

On a preceding page Garry Taylor describes the devastating effects that the influenza virus can have on human populations. A reminder that domestic fowl are just as — if not more — prone to flu comes in *The Veterinary Record* (**132**, 535–536; 1993). D. J. Alexander and colleagues describe an outbreak of infection in Norfolk, eastern England, in which almost 90 per cent of turkeys in a flock of 8,000 died in the space of a week; the outbreak occurred in 1991, ironically just before Christmas. The culprit was eventually identified as a highly pathogenic influenza A virus of subtype H5N1. Curiously, it seems that surviving birds were also infected, but by a less severe virus subtype. The authors speculate that either the two forms were present to start with, or that one underwent mutation into the other during the course of infection.

Wrapped up

USING a form of molecular origami, C.-H. Lung *et al.* have succeeded in synthesizing a long-sought and geometrically peculiar hydrocarbon cage compound (*Angew. Chem. Int. Ed. Engl.* **32**, 559–561; 1993). The compound, $p\text{-}[3^2,5^6]\text{octahedrane}(\text{CH})_{12}$, is made by folding up a cup-shaped precursor in an elegant application of classical methods. The final carbon framework, which includes six pentagonal faces and two triangular ones, is one of a family of structures first predicted in 1977, but of which only one, dodecahedrane $(\text{CH})_{20}$, has previously been made. Interest in such cage-like hydrocarbons stems largely from their rigid framework, to which all sorts of functional groups can be attached with precisely known geometry. As the cages also attach easily to cell walls, this opens the way to carefully designed biochemical probes.

after around 5000 BC¹⁰. The delta, therefore, became available for agricultural settlement at just about the right time. Other parts of northern Egypt may have been cultivated first — the oldest Neolithic sites of Lower Egypt, in the Fayoum, date back to about 5200 BC² — but in Upper Egypt, only the Badari region near Asyut had an agricultural tradition before 4000 BC¹¹. (A single tooth fragment considered to be of domestic cattle found at El Salamuni, dated to about 4600 BC¹², may indicate an older cultivated site in Upper Egypt.)

While groups from the eastern Sahara may have abandoned their desert homes and settled in the Nile valley as the climate became more arid, domestic wheat and barley were introduced into the Nile

valley from the Near East presumably through the now more amenable delta. Some domestic livestock may have arrived directly from the Levant along with the cereals, but it is also likely that the migrating Western Desert pastoralists brought their animals with them.

Stanley and Warne's work raises the intriguing possibility that, had conditions in the delta been favourable from the start of the Holocene, agricultural life might have taken root several millennia earlier. Egypt's Predynastic path to civilization might then have taken a very different course. □

D. L. Holmes is at the Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK.

1. Stanley, D. J. & Warne, A. G. *Nature* **363**, 435–438 (1993).

2. Hassan, F. A. *Afr. archaeol. Rev.* **3**, 95–115 (1985).

3. Krzyzaniak, L. in *The Nile Delta in Transition: 4th–3rd Millennium BC* (ed. van den Brink, E. C. M.) 151–155 (van den Brink, Tel Aviv, 1992).

4. Wendorf, F., Schild, R. & Close, A. E. (eds) *Cattle-Keeper of the Eastern Sahara: The Neolithic of Bir Kiseiba* (Southern Methodist Univ. Press, Dallas, 1984).

5. Neumann, K. in *Forschungen zur Umweltgeschichte der Ostsahara* (ed. Kuper, R.) 13–181 (Heinrich-Barth Inst., Cologne, 1989).

6. Wendorf, F. *et al.* *Nature* **359**, 721–724 (1992).

7. Gautier, A. in *Prehistory of Arid North Africa* (ed. Close,

A. E.) 163–187 (Southern Methodist Univ. Press, Dallas, 1987).

8. Holmes, D. L. *The Predynastic Lithic Industries of Upper Egypt* (BAR International Series 469, Oxford, 1989).

9. Hassan, F. A. *Norweg. Archaeol. Rev.* **19**, 63–76 (1986).

10. Kuper, R. Paper at *Int. Symp. Interregional Contacts in the Later Prehistory of Northeastern Africa* (Dymaczewo, Poland, 8–12 September 1992).

11. Holmes, D. L. in *The Nile Delta in Transition: 4th–3rd Millennium BC* (ed. van den Brink, E. C. M.) 301–316 (van den Brink, Tel Aviv, 1992).

12. Vermeersch, P. M. *et al.* in *The Followers of Horus: Studies dedicated to M. A. Hoffman* (eds Friedman, R. & Adams, B.) 163–172 (Oxbow, Oxford, 1992).

CLIMATE CYCLES

Cold answers to hot issues

David A. Peel

RICH archives of the Earth's climate have been kept in cold storage for more than 200,000 years in the ice sheets of Greenland and Antarctica. First results from two deep ice cores recently drilled in central Greenland show that geochemists are now able to detect climate shifts over periods as short as a few years in records sometimes several tens of thousand years old. Indeed, participants at a meeting in March* discovered that there are still surprises to be found in the relatively recent past. It seems that the 1908 meteorite impact at Tunguska in Siberia may have left its mark as an ammonium spike, the largest this century, in the Greenland ice sheet.

Not that the meeting was dedicated to such ephemera. Its primary purpose was to bring together the glaciologists whose ice cores contain the chemical evidence, and the atmospheric chemists and modellers who need to know what these traces reveal about crucial aspects of the atmosphere, including its circulation, turbidity and cloud cover, and oxidizing capacity. Until now, ice cores have been most valued for the information they offer on temperature (through stable isotope ratios) and on trace gases such as CO₂ and

CH₄, where the interpretation is most straightforward. Other constituents could potentially tell us about a much wider range of atmospheric characteristics, but the multiple sources involved, and the effects of transport and deposition processes, complicate the interpretation.

As the ice in the area of Summit in central Greenland is both ancient and unusually stable, the two deep cores recently drilled there (GRIP and GISP2) promise to give climatologists unprecedented detail on the Earth's history. The recent spike in ammonium formate, one of several sporadic concentration blips seen in the cores, was discovered by M. Legrand (Laboratoire de Glaciologie et Géophysique de l'Environnement, Grenoble). It may be coincidence that this occurred in 1908, the year of the Tunguska meteorite; but in the same snow stratum, H. Cachier (Centre des Faibles Radioactivités, Gif sur Yvette) found unequivocal evidence of solid aerosols due to carbonaceous combustion, confirming that, if not Tunguska, some exceptionally destructive event happened that year.

The ammonium peak in 1908 in fact bucked an unexpected trend for this century. Some environmentalists estimate that biomass burning has increased by 50 per cent this century and that ammonium

* *Ice Core Studies of Global Biogeochemical Cycles*, NATO Advanced Research Workshop, Annecy, 26–31 March 1993.

emissions globally may have increased threefold, so ammonium levels in the ice sheet should have followed a similar trend. Instead, the ice-core data seem to reveal a shallow decrease in recent decades. Atmospheric chemists have an answer for this: the true trend in emissions may have been masked by an increase in anthropogenic sulphate aerosol, which could scavenge the ammonium whilst it was being transported to Greenland. Indeed, new data from both the GISP2 and GRIP cores show that prominent increases in non-sea-salt chloride during the past century may also be linked to the recent acidification of the atmosphere, as sulphuric acid reacts with sodium chloride. Such interactions illustrate the dangers of simplistic interpretations of some of the more active constituents in ice cores.

One of the most significant results discussed at the meeting was the revelation that methane levels measured in the Summit core closely paralleled the inferred temperature changes during periods of extremely unstable climate towards the end of the last glacial period (from 40,000 to 20,000 years ago), as well as during the 'Younger Dryas' cold period and during the 'Little Ice Age' of about AD 1550–1800 (B. Stauffer, Bern University). This is the strongest evidence yet for a positive climate forcing by one of the main greenhouse gases (molecule for molecule, methane is 21 times as powerful a greenhouse gas as carbon dioxide).

Concentrations of methane in the atmosphere are limited by photochemically derived OH radicals, which are the principal oxidizing agent in the troposphere and limit the lifetime of many pollutants in the modern atmosphere. Recent photochemical model calculations (A. Thompson, NASA/Goddard Laboratory) predict methane mixing ratios that are broadly confirmed by the ice-core data for both the Last Glacial Maximum and for the pre-industrial Holocene, and suggest that the changes are largely driven by changes in emission rates.

Most of the other species of interest have very much shorter residence times in the atmosphere and so show much greater geographical variations. The most obvious difference between the Earth's polar regions is that the Arctic regions are landlocked, surrounded by the great industrialized nations, whereas Antarctica is surrounded by the Southern Ocean. This marked asymmetry can help to separate the contributions from different sources, especially for the tropospheric aerosol components which have a particularly short residence time in the atmosphere. One consequence is that it is still possible to study in recent snows the natural biogeochemical cycles of nitrogen and sulphur in Antarctica, where the effects of pollution are virtually undetect-

able; in the Arctic, these cycles have been strongly perturbed by pollution.

Sulphate aerosols are thought to influence climate feedback both by scattering solar radiation directly and by altering the reflectivity of clouds. There may also be feedback between climate and the natural sulphur cycle. Given the marine setting, it is not surprising that sulphur in Antarctic ice is normally dominated by compounds derived from marine algae. One of these, methanesulphonic acid, varies strongly on both decadal and glacial to interglacial timescales, suggesting important, large-scale changes in marine biological activity. Although this may well be so, recent work has shown that wind speed (patently a climatic factor) strongly affects the rate of exchange of the marine sulphur compounds into the atmosphere (E. Saltzman, Miami University).

Moreover, the production of dimethyl sulphide, the precursor of methanesulphonic acid, depends on the distribution of planktonic species. The variations observed in the ice-core record are thus the net result of several complex and interconnected factors. Patterns of behaviour in the Arctic and the Antarctic seem to be strikingly different during the coldest parts of the last ice age. Whereas methanesulphonic acid shows large increases at several Antarctic sites, new data from Renland, Greenland, a low-altitude site, show it to decrease (Saltzman). Possibly this reflects changes in a local source for the northern data. Research on these sulphur species should have much to add to our knowledge of changes in ocean conditions, a further crucial factor in climate change.

The behaviour of sulphur is complex because so many varied sources are involved, including the overwhelming anthropogenic contribution in the Northern Hemisphere during the past few decades. In the Arctic, progress is being made towards establishing the links between climate and aerosol chemistry, including the sulphur species (L. Barrie, Atmospheric Environment Service, Ontario). This research is done at ground-based stations such as Alert; if it could be extended to areas around ice-drilling sites in Antarctica and Greenland, where it could be combined with investigations of how atmospheric impurities are deposited in the local snowfall, we should obtain a firmer basis for translating ice-core chemistry into atmospheric chemistry. Ice cores could then offer a relatively low-cost means of extending direct atmospheric sampling to the distant past, where no satellite probe can reach, providing urgently needed information for global models of atmospheric chemistry. □

David A. Peel is at the British Antarctic Survey, NERC, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

DAEDALUS

Waxing Moon

A RECENT Russian satellite reflected back to Earth a feeble artificial moonlight. Daedalus now plans to improve the real thing. He points out that the Moon reflects only about 5 per cent of sunlight — little better than soot. If, like magnesium oxide, it reflected 95 per cent of sunlight, moonlight would be twenty times as bright.

At first Daedalus had visions of some desperate Apollo project, with rocket loads of magnesium oxide being ferried to the Moon, and astronauts with bulldozers spreading it about. But he now has a much neater idea. A rocket motor burning magnesium in oxygen could eject an exhaust plume of magnesium oxide at more than 5 km s^{-1} . Fired in the vacuum of space, this plume would travel quite ballistically. It could be aimed at the Moon from a great distance, and would impact it as evenly distributed dust. Magnesium oxide smoke is so finely divided that a mere ten million tonnes of it could cover the whole visible face of the Moon with a dazzlingly white reflecting layer.

The obvious technology for this job is the high-velocity gun developed as a space launcher for the Canadian 'Harp' project. It should be set up in the Himalayas, perhaps on the already overcrowded Mount Everest, to minimize air-resistance to the projectile. Once in space, this would be accelerated towards the Moon by a booster rocket. At its closest approach, it would ignite a motor burning magnesium powder in lithium perchlorate. This mixture burns to magnesium oxide and lithium chloride (another very white solid). An orientation package would aim the fast-moving plume so as to impact the lunar surface. Guns are so cheap compared to space rockets that a heavy lunar barrage could be sustained for years, until the whole surface of the Moon was evenly dusted with magnesium oxide.

Life on Earth would be transformed. Dazzlingly bright moonlight would save billions on street lighting everywhere. Lovers would blink in its unromantic glare, while nocturnal criminals and conspirators vainly sought concealment from the lunar searchlight. Siberia and the Canadian north, released from their dark winters, would flourish as never before. Flowering plants might be confused by the seemingly endless summer daylight; but the monthly new-Moon period should give them an adequate clue to the true season. The cost of the project would be recovered many times over. But to fund its early days, the first white strokes on the dark lunar surface should perhaps take the form of advertising slogans. David Jones

Which home for coelacanth?

SIR — In 1938, the first living coelacanth (*Latimeria chalumnae*) was trawled off the sandy estuary of the Chalumna River on the east coast of South Africa at a depth of 75 m. All 200 catches since then have come from the Comores, northwest of Madagascar. In August 1991, a gravid female was trawled during the day at a depth of 40–45 m off Quelimane, Mozambique (Fig. 1). This raised the question of whether other coelacanth populations exist along the African mainland or whether coelacanths caught off the African coast are strays.

Our submersible survey at Grande Comore and Anjouan reveals that *Latimeria* is a bottom-oriented drift hunter inhabiting the relatively calm waters of the rocky slopes at around 200 m depth. Being inactive during the day, it retreats to the still waters of lava caves¹. By contrast with the Comorean habitats, the South African and Mozambique locations are exposed to the current, have flat, sandy bottoms and are shallow, seeming unsuitable for coelacanths.

To test the assumption that the Mozambique coelacanth does not belong to the Comorean population, we compared mitochondrial DNA sequences from the most rapidly evolving part of the control region as well as band-sharing frequencies of multilocus DNA fingerprints of several individuals from the Comores and of the Mozambique specimen. Only two different sequences were obtained, varying by one transition at one position in 261 base pairs of the control region among 16 individuals of the Comores. The Mozambique sequence was identical to one of them. This result strongly suggests that the Mozambique and the Comorean coelacanths belong to the same population.

We used DNA multilocus fingerprint-

ing to exclude very recent population differentiation. DNA samples of sufficiently high molecular mass (>30 kb) were digested with *Hinf*I, *Hae*III and *Alu*I, separated on 0.8% agarose gels and hybridized to a panel of simple repeat oligonucleotide probes (see ref. 2). Several of these yielded a uniform multilocus hybridization pattern whereas others produced uninformative smear signals using *Latimeria* DNA as target. Most informative with respect to individualization were (CT)₄ (CA)₅ and (CA)₈ among more than 20 simple repeat probes tested.

Five different individuals from Grande Comore were compared with late-term pups of the gravid female from Mozambique. Determination of the band-sharing index revealed a high similarity of the fingerprints from the Comorean fish, showing a close relationship and low genetic variability comparable to that in close-bred stocks of laboratory fish³. The Mozambique samples are as similar to each of the Grande Comore fish as they are to each other (Fig. 2).

If a separate population of coelacanths exists at the East African coast, we would have expected substantially different fingerprint patterns from the Comorean population. Thus, mitochondrial as well as nuclear markers indicate that the Mozambique coelacanth originated in the Comorean population. An alternative explanation would be that there is an African coastal population in continuous genetic contact with the Comores. But this seems very unlikely.

The Comores are isolated volcanic seamounts surrounded by depths of more than 3,000 m. Coelacanths are adapted to temperate waters of the twilight zone¹. Therefore, benthic migrations through cold waters of the deep sea are extremely unlikely and so the animals could have migrated only by passive drift through the pelagic. However, during almost 200 dives, we never found coelacanths in open waters; drifting in the pelagic zone must be extremely rare. Therefore we believe that the Mozambique coelacanth is a stray originating from the Comores. The Comorean archipelago is exposed to the strong southwest-flowing Mozambique current⁴ which could have transported the fish (Fig. 1).

Although we have not yet been able to obtain

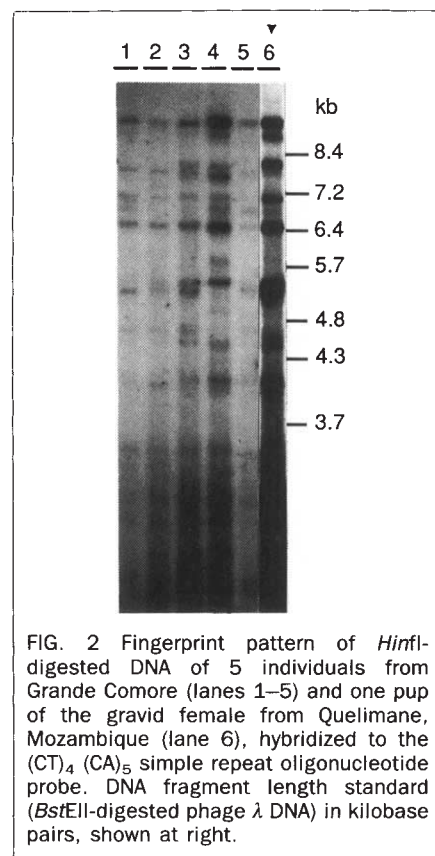


FIG. 2 Fingerprint pattern of *Hinf*I-digested DNA of 5 individuals from Grande Comore (lanes 1–5) and one pup of the gravid female from Quelimane, Mozambique (lane 6), hybridized to the (CT)₄ (CA)₅ simple repeat oligonucleotide probe. DNA fragment length standard (*Bst*EII-digested phage λ DNA) in kilobase pairs, shown at right.

polymerase chain reaction amplifications of the formalin-fixed South African coelacanth, this specimen could also have been a stray. This has already been suggested by J. L. B. Smith, the discoverer of the living coelacanth. The South African location is exposed to the Agulhas current, in part a continuation of the Mozambique current. Therefore the South African coelacanth could also have been derived from the Comores, although we cannot exclude an origin from the east coast of Madagascar.

U. Schiewen

H. Fricke

Max-Planck-Institut für
Verhaltensphysiologie,
8130 Seewiesen, Germany

M. Scharl

Physiologische Chemie I,
Theodor-Boveri-Institut für Biowissen-
schaften der Universität Würzburg,
8700 Würzburg, Germany

J. T. Epplen

Molekulare Humangenetik,
Ruhr-Universität,
4630 Bochum, Germany

S. Pääbo

Zoologisches Institut der
Universität München,
8000 München, Germany

1. Fricke, H. et al. *Envir. Biol. Fish.* **32**, 287–300 (1991).

2. Scharl, M. *Genetics* **119**, 679–685 (1988).

3. Nanda, I., Scharl, M., Feichtinger, W., Epplen, J. T. & Schmid, M. *Chromosoma* **101**, 301–310 (1992).

4. Donguy, J. R. & Piton, B. *Oceanol. Acta* **14**(6), 549–558 (1991).

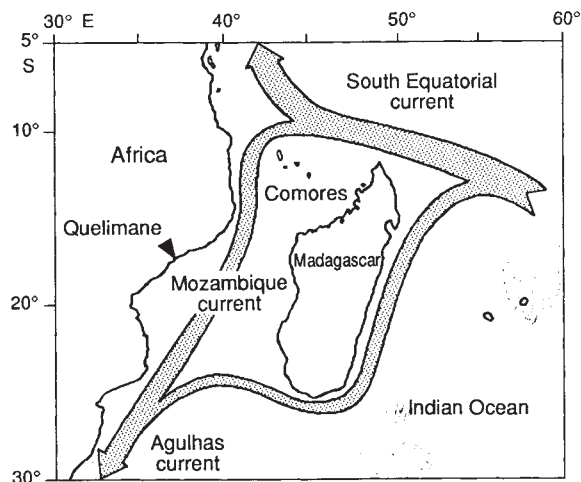


FIG. 1 Classical model of current distribution in the Mozambique Channel⁴. Arrowhead, location of the second coelacanth catch outside the Comores, at Quelimane.

Signalling in planktonic larvae

SIR — Berridge¹ points out that cyclic AMP and inositol trisphosphate messenger pathways may interact to enhance odour discrimination. Readers may be interested to learn that a similar interaction can apparently enhance chemosensory responsiveness to chemical signals recognized as morphogens by planktonic marine invertebrate larvae in the ocean.

Many benthic marine invertebrates reproduce to yield large numbers of small larvae that are dispersed in the plankton. In several species, substratum-specific settlement of larvae from the plankton and metamorphosis are induced by chemosensory recognition of exogenous chemical signals associated with specific recruiting surfaces. Site-specific settlement and metamorphosis of *Haliotis rufescens* (red abalone; gastropod mollusc) larvae are regulated by two convergent chemosensory pathways^{2,3}. The inducer recognized by these larvae is a GABA-mimetic peptide produced by the algae on which the larvae are induced to settle⁴; binding of the morphogenetic signal molecule by externally accessible chemosensory receptors is apparently transduced by a cyclic AMP- and calcium-dependent pathway leading to the induction of settlement and metamorphosis^{3,5}.

Lysine dissolved in sea water acts through a separate pathway to increase post-receptor responsiveness of the larvae to the required morphogenetic signal by as much as 100-fold. The binding of lysine to its receptor is transduced by a receptor-associated G protein, which in turn controls the activation of a diacylglycerol-stimulated, calcium-stimulated protein kinase C (refs 2, 3, 6). This interaction between the morphogenetic (cyclic AMP-dependent) and regulatory (inositol trisphosphate/diacylglycerol-dependent) chemosensory pathways, by increasing the responsiveness of the larvae to their morphogenetic cue when threshold concentrations of facilitating amino acid are detected in the water column, may thus increase the propensity of larvae to settle on the recruiting algae in potentially favourable (for example, nutrient-rich) areas^{2,3,5}. Such interactions between larval signal-

ling pathways, controlled by different cues from the environment, could provide a capacity for fine-tuning the recognition and selection of habitats favourable for metamorphosis, and could help explain the temporal and spatial variability in recruitment of some marine species.

Daniel E. Morse

Marine Biotechnology Center,
University of California,
Santa Barbara, California 93106, USA

1. Berridge, M. J. *Nature* **361**, 315–325 (1993).
2. Baxter, G. & Morse, D. E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1867–1870 (1987).
3. Morse, D. E. in *Abalone of the World* (eds Shepherd, S. A. et al.) 107–119 (Blackwell, Oxford, 1992).
4. Morse, A. N. C. *Systematics Ass. Spec. Vol.* **46**, 385–403 (1992).
5. Tapido-Rosenthal, H. G. & Morse, D. E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7658–7662 (1986).
6. Baxter, G. T. & Morse, D. M. *Biol. Bull.* **183**, 147–154 (1992).

Not the earliest solar eclipse

SIR — Semitists and historians of astronomy have devoted much discussion to the Ugaritic text RS 12.061 (see ref. 1), generally interpreting it as an account of a solar eclipse, thought to have occurred in the fourteenth or the thirteenth century BC (refs 2–5). In the course of preparing a new edition of the Ugaritic texts dealing with various aspects of ritual, we have concluded that the text cannot refer to a solar eclipse.

The text and our translation are shown in the box.

(1) b tt . ym . hdt	During the six days of the (rituals of) the new moon of
(2) h̄yr . 'rbt	(the month of) <i>Ḫiyyāru</i> , the sun
(3) špš [.] tgrh	set, her gatekeeper (being)
(4) r̄sp	<i>Rashap</i> .
<i>Verso</i>	
(5) [w á] dm [.] tbqrn	The men (?) shall seek out
(6) skn	the prefect.

Ugaritic grammar indicates that *b tt ym hdt h̄yr* should not mean “on the sixth day of the month of *Ḫiyyāru*” (which would be *b yrh h̄yr b ym tdt*), but “during the (first) six days of the new moon (rituals) of the month of *Ḫiyyāru*”. This was generally perceived by de Jong and van Soldt^{2,3}, who noted that *tt*, if a number, must be cardinal, but who chose on no explicit basis to interpret *b tt* as referring to the hour of the day. They interpreted the text as referring to a solar eclipse, though they provided no new philological data to support that interpretation. Indeed, there is no evidence from any of the ancient Semitic languages for the verb ‘*rb*, literally “to enter”, having the meaning “to be eclipsed” with “sun” as its subject. The idiom consisting of ‘*rb* plus “sun” is rather the standard idiom for the

Sun setting⁴. A relic is preserved in Hebrew where the verb ‘*rb* has been replaced by *bw*’, but the noun ‘*ereb*, “evening”, has been retained.

The simple meaning of *b tt ym hdt* and of ‘*rbt špš* is therefore that the Sun “set” on the first 6 days of the month in question. This fact in itself being of no significance, the reason for the writing in clay of this text appears to lie in the next clause, according to which the deity *Rašap* was the Sun’s gatekeeper as she entered the underworld. *Rašap* is known to be one of the primary deities of the underworld, the West Semitic equivalent of Mesopotamian *Nergal*. The basic question regarding the interpretation of this text is to identify this deity serving as the Sun’s gatekeeper. The only solution for which there are presently any data of which we are aware is that which identifies *Nergal/Rašap* with *Mars*^{2–4}.

We interpret the text as meaning that Mars was visible at the setting of the Sun during the first 6 days after first visibility of the new Moon in the month of *Ḫiyyāru* in an unstated year (a dating system that included the notation of the year is not yet attested at Ugarit), after which the planet was no longer visible. According to this hypothesis, the planet Mars would have appeared in the west with the lunar crescent on the first day of *Ḫiyyāru*, and would have been visible for each succeeding evening until the seventh, when it was not visible. The heliacal setting would, therefore, have occurred on the sixth day. Because within the period ~1400–1200 BC there are about five heliacal settings of Mars in the part of the year corresponding

to the month of *Ḫiyyāru*, the text cannot be uniquely dated.

Thus our hypothesis, which insists on the 6-day duration of the observed event, is significantly different from those that consider the possibility of the text referring simply to the heliacal setting of Mars (for example refs 2–4).

D. Pardee

The Oriental Institute,
N. Swerdlow
Departments of Astronomy and
Astrophysics and History,
University of Chicago,
Chicago, Illinois 60637, USA

1. Virolleaud, C. *Syria* **28**, 25–27 (1951).
2. de Jong, T. & van Soldt, W. H. *Jaarber. Ex Oriente Lux* **30**, 65–77 (1987–88).
3. de Jong, T. & van Soldt, W. H. *Nature* **338**, 238 (1989).
4. Walker, C. B. F. *Nature* **338**, 204–205 (1989).
5. Mostert, R. *Nature* **341**, 112–113 (1989).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Culture shock

David DeVorkin

Inside NASA: High Technology and Organizational Change in the American Space Program. By Howard E. McCurdy. Johns Hopkins University Press: 1993. Pp. 215. \$32.95, £27.50.

WHEN government funding of institutions is cut back after a crisis or because of changing national policies or imperatives, what, if anything, do the institutions do to survive? Do they innovate and streamline through creative restructuring? Do they become rigidly complex bureaucracies intent on survival? Or do they simply fade away? And as these institutions fight for survival, how do they change? These are the questions raised by this important book.

Howard McCurdy, professor of political science and public affairs at American University in Washington DC, has analysed the changes in NASA (National Aeronautics and Space Administration) during its first three decades. He finds that as NASA cut back on the size of its operations in the early 1970s, the number of administrative staff at headquarters rose relative to the number of engineers and scientists; that a growing preference for contract management suppressed in-house research and development; and that a vigorous culture of testing and verification gave way to an ageing management preoccupied with survival and avoiding risks. These and other shifts, McCurdy contends, were most noticeable during the second and third decades of the agency, and meant that programmes and projects were chosen more to keep people in employment than to address issues relevant to national needs. He argues that although NASA tried to diversify in the post-Apollo era, its organizational responses to a weakened national mandate only added to its rigidity and inability to adapt.

McCurdy's general conclusion is that high-technology institutions can exist as government research and development agencies only temporarily, when their missions are crystal clear and politically essential. To make his point, he briefly alludes to the fates of the Depression-era Tennessee Valley Authority, the wartime Radiation Laboratory at the Massachusetts Institute of Technology and the Manhattan Engineer District (the US atomic bomb project), drawing on the growing amount of literature on organizational culture and institutional histories. Such agencies could not enter the govern-

ment "mainstream and sustain high-performance cultures for extended periods of time". In fact, McCurdy sees NASA as a "textbook example" of the life of a high-technology organization from its inception, rapid growth and bureaucratization, through to its decay.

McCurdy has relied on dozens of interviews and an extensive questionnaire completed by about 700 first- and second-

programme, officials from the Air Force Ballistic Missile Program worked to strengthen project management within NASA. The resulting creative tension provided the necessary but temporary balance, McCurdy argues, between independent technical expertise and strong management. This "loose-tight" combination, according to McCurdy, is what made the Apollo project work, just as it did in other successful large-scale, high-technology projects. But the unstable equilibrium did not last long; it started to shudder during the last phases of the Apollo project, and collapsed soon after.

McCurdy's analysis fits the available histories reasonably well. He roots out the causes of change, interpreting how NASA developed as an organization during its expansionist Apollo years and how it tried

to satisfy its passion for manned space flight and the permanent human occupation of space at a time when the US government decided that it had better ways to spend its money.

It is an impressive and thought-provoking study, but there are problems with some of the details and methodology. First, interviews are delicate instruments. It is not surprising that first-generation NASA employees look back with pride on the early days of the agency, but with bewilderment on how their successors and — as some argue — their nation have squandered their legacy. McCurdy is at first over-careful not to dispute the testimonies of his interviewees. Only in the second half of the book does he introduce any counter-evidence to allow readers to put the oral histories in context.

Interviews provide a useful reflection of self-image and attitude, but they are not, on their own, good sources of fact. In this respect, McCurdy might also have been more critical of earlier written sources that often similarly rely on oral testimonies. Finally, although outright blunders are rare, one stands out. He understates by a factor of 1,000 the cost of the Orbiting Astronomical Observatory and Ranger series. This weakens his point that the real cost of later missions such as the Hubble Space Telescope increased as their complexity grew. The actual increases are far less dramatic.

But McCurdy is surely on the right track. His valuable book makes the literature on organizational cultures accessible and reveals new ways to look at high-technology agencies. □

David DeVorkin is at the National Air and Space Museum, Smithsonian Institution, Washington, DC 20560, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Flight for survival — launch of the Apollo 9 mission in 1969.

generation NASA professional employees (administrators, scientists and engineers). Using NASA's official histories and the standard (and still inadequate) literature on space history, he shows that the early NASA was a blend of many different cultures: Wernher von Braun's German-dominated group in the Army Ballistic Missile Agency at Huntsville; the laboratories of the National Advisory Committee for Aeronautics; rocketry groups from both the Naval Research Laboratory and the California Institute of Technology's Army contract operation at the Jet Propulsion Laboratory; and management specialists from the Air Force Ballistic Missile Program. The first three formed NASA in its first two years and fostered a strong emphasis on hands-on testing and verification. When the Apollo project was approved as a crash pro-

Problems in evolution

Mark Ridley

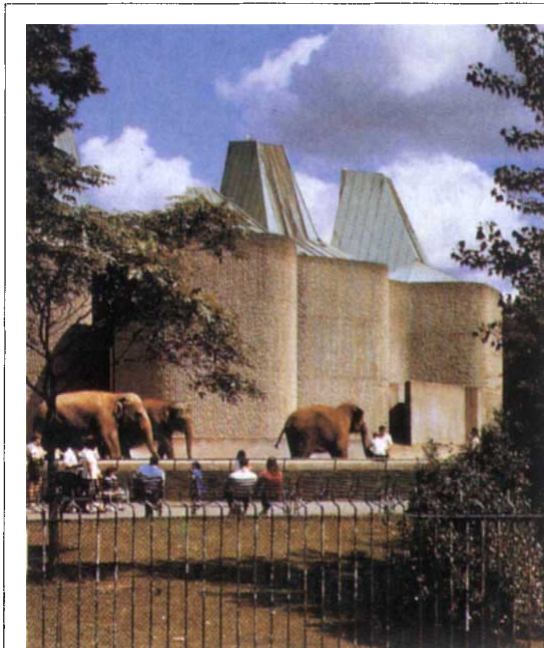
Natural Selection: Domains, Levels, and Challenges. By George C. Williams. Oxford University Press: 1992. Pp. 208. £35, \$55 (hbk); £14.95, \$24.95 (pbk).

WITH *Adaptation and Natural Selection* (1966) and *Sex and Evolution* (1975), G. C. Williams released a pair of books that have influenced evolutionary biology in the late twentieth century perhaps more than any other pair of books by any other biologist. *Adaptation and Natural Selection* destroyed the theory of group selection so comprehensively that woolly headed group-selective arguments have vanished from the professional literature. In *Sex and Evolution* he argued so persuasively that sexual reproduction is the outstanding evolutionary puzzle that evolutionary biologists now discuss sex more than perhaps any other topic.

Williams's new book is less monolithic than either of its predecessors, and combines the argumentative style of both. It contains a series of reflections on controversial topics in modern evolutionary biology, together with suggestions about some further unsolved problems that ought to be receiving more attention. It ranges widely, and many kinds of specialist could sample it for inspiration and fresh thinking. Only the future can reveal whether it will have as much influence as his previous books, but it does contain several ideas that are big enough to make it a possibility.

Williams's explanation for "stasis" (that is, fossil lineages that do not change in form for long periods) is probably the biggest new idea in the book. There are three main existing ideas, and he disagrees with them all. S. J. Gould has suggested that species have "genetic and developmental coherences that resist selective pressures"; but Williams reasons, from microevolutionary evidence, that this "is clearly not true". Others have suggested that species stay the same because, when conditions change, they migrate; but Williams dismisses this as a "fable [that] will not bear close examination" — the selective forces almost certainly alter when the biogeographic range of a species shifts a few thousand miles. Williams also doubts whether stabilizing selection, at least in its simplest form, is the reason because there is so much evidence of rapid change in modern populations.

So why is the form of fossils so often constant through time? Consider the evolutionary pattern of the modern three-spined stickleback. In North America, it is primarily a coastal species, but innumer-



LONDON Zoo's architecture is an exhibit in its own right. Building and rebuilding has taken place almost continuously since 1827, revealing as much about the history of design and construction as that of animal display. The wittily zoomorphic Elephant and Rhino Pavilion of 1962–5, pictured here, was one of a succession of large buildings erected between 1958 and 1976. As an exercise in New Brutalism, it broke with international orthodoxy in zoo architecture, combining "the fantastic with the functional". This photograph is taken from *The Buildings of London Zoo* by Peter Guillery (Royal Commission on the Historical Monuments of England, £12.95 (pbk)).

able forms have evolved as the fish have migrated upstream into individual freshwater tributaries. Williams suggests these derived forms may be evolutionarily short-lived, as their niches will soon disappear; but the main coastal niche, and its ancestral form of the stickleback, will persist for longer. An imaginary series of fossil samples through time would be most likely to draw out the unchanging ancestral lineage, as it occupies its enduring, and little changing, niche. "The appearance of stasis in the fossil record would result from an enormous variability in the persistence of ecological niches."

The explanation is an instance of what Williams prefers to call "clade selection": the process by which different lineages survive better or worse through evolutionary time. It is more familiar as "species selection", but Williams dislikes that expression because (he suggests) the process is no more typical of species than of lineages (or clades) at higher and lower taxonomic levels. In the exemplary sticklebacks, clade selection would be operating among lineages within one (at least conventionally recognized) species. Williams also rejects any connection between clade selection and the theory of punctuated equilibrium, and argues that evolutionary change is not particularly concentrated in speciation events (indeed he concludes that "speciation in the usual [Mayrian] sense has no special significance for macroevolution"), and that "the peripheral isolate theory would seem to be of little use". But the general thesis of the book should warm the hearts of palaeobiologists: "the microevolutionary process that adequately describes evolution in a population is an utterly inadequate account of the Earth's biota". The uncoupling of microevolution and macro-

evolution in Williams's theory is not, as in the orthodox view, due to speciation, but to that "enormous variability in the persistence of ecological niches". Much of the microevolution we observe takes place in ephemeral niches, whereas the enduring niches determine evolution on the grand scale.

Williams discusses many other questions. He argues that "Haldane's dilemma" (the problem of genetic load) was never solved, even though interest in it has faded away; he raises questions about unobserved adaptations that might reasonably be expected — viviparous turtles, more facultative sex determination, more flexibility of metabolic temperatures; and has a wonderful section in which he argues that female pheromones in moths are not sexual signals at all. There are short asides, about topics such as cultural evolution or play, that I suspect would be better places to start new research than the whole monographs that have been written on them. There are so many ideas that no reader will agree with them all. Thus I am sure the question Williams raises about the comparative method really *has* been answered, and I believe his criticisms of M. T. Ghiselin and D. L. Hull's species-as-individual concept are beside the point. But these are small matters in a delightful book. As with any frontier investigation, it will attract some scoffers — the citation police, the factually fixated, anal-retentive equation-processors, and dullards and philistines generally. But they can all be ignored. And — who knows? — the book could free some of them from their vices. □

Mark Ridley is in the Departments of Anthropology and Biology, Emory University, Atlanta, Georgia 30322, USA.

Growth and form

Jane Maienschein

Styles of Scientific Thought: The German Genetics Community, 1900–1933. By J. Harwood. *University of Chicago Press*: 1993. Pp. 423. \$74.75, £51.95 (hbk); \$27.50, £17.95 (pbk).

GENETICISTS, Jonathan Harwood contends, showed two different predominant styles of thought between the First and Second World Wars. Indeed, he suggests that these styles characterize German scientists more generally and may extend beyond Germany because they represent alternative responses to modernization. They are most characteristically German because of the particular German social, political and institutional conditions that constrain and define the styles, yet similar features may be found elsewhere.

Harwood's first style is the "comprehensive". Comprehensive geneticists tend to think in larger terms: they tend to embrace more holistic rather than more atomistic theories and to pursue more comprehensive problems, including evolution and development alongside heredity. They tend to come from more 'mandarin' backgrounds, with an emphasis on culture and unity of thought and values, and they remain sceptical about the focus on industry and application brought about by modernization. Although not more likely to be conservative rather than liberal politically, they may be more likely to be politically active. And they are more likely to be explicitly credited in obituaries with the breadth of their concerns.

Contrasts

"Pragmatic" geneticists, on the other hand, tend to think and work more narrowly, with a greater focus on the practical and productive. They tend to think in terms of parts rather than wholes, and they rarely consider development or evolution when they work on problems of heredity. They come more often from 'modern' backgrounds, with less well established families and training in more modern schools. Culture and the arts are less important to them, as they focus more exclusively on the work at hand and embrace a practical emphasis on the process of modernization. They are less likely to be politically active, and they remain 'outsiders' to the political community. Obituaries typically cite their important research and its applications and thus emphasize the depth of their interests.

In addition, pragmatics are more often outside the established universities, perhaps residing in agricultural or other

applied research stations, whereas comprehensives serve as keepers of tradition within the universities. It might seem that the type of institution, which correlates highly with the type of problems selected in genetics, determines the problems. Yet perhaps the set of values is already established by social and other factors and directs the individual to an appropriate institution, which may in turn reinforce the chosen style. As Harwood points out, it is plausible that "Class proposes, institution disposes".

Perhaps also there are national differences in style, in part for social or class reasons and in part for institutional reasons. Harwood finds the contrast between German and American genetics instructive, yet acknowledges that it does not warrant the conclusion that neatly defined national styles are involved. Most historians have concentrated on American and British contributions to genetics as if these told the story of genetics generally. But the Americans may actually have followed an atypical path, Harwood suggests (although it is not clear given his perspective why any path should be thought of as either typical or atypical). They were, he says, "simply not interested in the theoretical issues" and hence missed the larger, more comprehensive picture. By training, institutional affiliation and metaphysical inclination, the Americans were more pragmatic. Germans who left for the United States found Americans generally to be narrow, uncultured, unconcerned with evolution and development and rather uninteresting.

For example, the Americans "almost universally" rejected the role of the cytoplasm in heredity. Because they concentrated on transmission genetics, they did not see the importance of Richard von Wettstein's theory that the plasmon is a regulatory mechanism in gene expression. Nor did they accept other forms of cytoplasmic inheritance or "persistent phenotypic changes" such as 'dauermodification', or the idea that the cytoplasm might serve as a substrate for genetic action. The fact that such ideas could contribute to interesting evolutionary theories had no appeal either, because neither evolution nor development was their chosen research problem.

Drawing on profiles compiled from selected groups, Harwood acknowledges that the numbers are necessarily small and the data incomplete. His sociological model cannot explain each of the individual variations, but it does allow for the existence of such differences.

Biographies will tell us more, and more research will probably bring revisions of Harwood's categories and model. He knows that. Indeed, he encourages others to enter the discussion and acknowledges that they will prob-

ably disagree on some points. He has offered such a wealth of ideas within such a richly argued framework that the contribution will stand as important even if some of the conclusions fall under the weight of additional evidence and alternative interpretations. Here are a few points likely to undergo revision.

Comparisons

The comparison between Germans and Americans needs more work. The suggestion that more Americans were like the German pragmatics and few like the German comprehensives seems right. The social, political and educational structures were sufficiently different, however, to make the reason and the details different. Also, Harwood cites T. H. Morgan as a clear pragmatic, even a "look-alike" for the German pragmatic exemplar Erwin Baur. Yet Morgan is also the one American geneticist whom the German *émigrés* considered cultured and as having broad interests. Harwood sees Morgan as a transmission geneticist, who therefore should not fit the comprehensive pattern, yet he does not discuss Morgan's deep and real developmental interests.

Second, because Harwood has had to rely on obituaries and limited additional evidence, his interpretations of the various individuals' views may be wrong. Because he offers generalizations from compiled cases, and because the sample size is limited, a few shifts could undercut the claims for generality of pattern. He may very well be wrong about the way in which various characteristics cluster. But at least he has provoked historians to look for such clusterings and has provided a promising starting point for doing so.

Third, while concentrating on problem choice, theory choice and some metaphysical commitments of his geneticists, Harwood has left out epistemology — or

1993 Science Book Prize

Steven Rose has won this year's £10,000 Science Book Prize for *The Making of Memory: From Molecules to Mind* (Bantam, 1992). The prize, sponsored by the French chemical and pharmaceutical company Rhône-Poulenc, is awarded annually for books judged to contribute most to the public understanding of science. For a review, see *Nature* 361, 127 (1993).

Highly commended were *Genius: Richard Feynman and Modern Physics* by James Gleick (Pantheon/Little, Brown) and *The Malaria Capers* by Robert S. Desowitz (Norton), respectively reviewed in *Nature* 360, 375 and 356, 300 (1992).

In next week's issue, Steven Rose assesses Simon LeVay's *The Sexual Brain*.

at least has not explicitly and consistently included important epistemological commitments. He refers to the Vienna Circle as moderns and cites their abhorrence of metaphysics, for example, but does not go on to explore the epistemological implication. Nor does he examine the role of changing and alternative sets of epistemological values for the Americans, who often adopted a pragmatic research programme because in that way they could get something that they could accept as reliable knowledge. Similarly, for some of the Germans epistemology was of central importance, but not for others. This too cries out for more study.

Finally, Harwood only touches on the exciting question of why researchers did *not* take up certain problems, approaches or theories. Often scientists make choices because they are attracted to a particular research programme; but perhaps they are constrained by their basic commitments to other values, their social background or their institutional setting. Harwood criticizes Ludwik Fleck and Gerald Holton for saying that a style or themata can "guide" scientists' work because "ideas of themselves do not coerce". Yet a deep commitment to ideas — however instilled — can coerce. Understanding where, how and why the guiding or coercing occurs can help us to understand not just that styles of scientific thought exist but how they influence scientific development.

Harwood has provided a wonderful stimulus to further thinking about the development of genetics. His book should provoke exciting debates and important insights, even if new work undercuts some of his own suggestions. □

Jane Maienschein is in the Departments of Philosophy and Zoology, Arizona State University, Tempe, Arizona 85287-2004, USA.

Ancient oceans

E. Irving

Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans. By Rob Van der Voo. Cambridge University Press: 1993. Pp. 411. £60, \$89.95.

ROB Van der Voo has devoted most of his life to the study of the palaeomagnetism of Palaeozoic rocks (590–248 million years old). From such studies, one may calculate the movements of the continents that have constrained the shape and distribution of oceans. Here he reviews the results as they relate to the evolution of the Palaeozoic Iapetus, the Late Palaeozoic to Early Mesozoic Tethys, and the Mesozoic and Cenozoic Atlantic Oceans.

A large obstacle in such studies is widespread remagnetization. In the Late Palaeozoic there was a global tectonic event that resulted in the final assemblage of Pangaea. Rocks were deformed, buried and flooded with basinal fluids to produce new magnetic minerals and stable magnetization. Older, more consolidated Precambrian rocks were little affected, but many Palaeozoic deposits were partially or completely remagnetized. Evidence for Late Palaeozoic remagnetization was discovered in the mid-1960s, but it was not until the early 1980s that its global nature was accepted. Meanwhile, workers had been fooled by this unwelcome phenomenon, mistaking younger for older magnetizations. Through all these vicissitudes, Van der Voo and others in the community have persisted; it is to their credit that palaeomagnetism is now as indispensable in the study of the tectonics of the Palaeozoic as it is in that of later eras. This book is a monument to their efforts. All tectonicists should read it.

The book's leitmotiv is the diagram of variation in declination and palaeolatitude. There are dozens of them. They make clear the general picture, but the errors in the measurements are not given: that is, errors critical to the estimation of drift rates and of the times of convergence and amalgamation of moving crustal blocks. Observed values are compared with those expected from neighbouring crustal blocks, and differences are interpreted as relative rotation about local vertical axes and relative palaeolatitudinal displacement. Calculations of this relative rotation and displacement and their errors are not generally made, and procedures for calculating them are not explained.

A second prominent subject is palaeomagnetic tests of continental reconstructions. Van der Voo provides a masterly demonstration of the superiority, over others, of Sir Edward Bullard *et al.*'s fit of the northern continents. His discussion of the reconstructions of the Gondwana continents is serviceable, but less mature because of the sparsity of data for the Southern Hemisphere (although he gives up too easily on Pangaea B). Future analyses will no doubt limit further options for mappers of the Palaeozoic Earth, and the book admirably sets the stage for their work.

There are many charming asides where the author reveals his habits of thought, one of which is to complain regularly about the frequent unsuitability of data for the task he has in mind. He forgets, perhaps, that some were obtained with different objectives. Data, good or bad, suitable or unsuitable, have an unfortunate habit of becoming out of date, and what really counts in the long run is the quality and durability of ideas developed from them. Far better to rejoice that, from a standing start in the late 1940s,

palaeomagnetists have contributed many fundamental ideas to Earth science, contributions to which this book bears undeniable witness. □

E. Irving is emeritus scientist in the Pacific Geoscience Center, Geological Survey of Canada, PO Box 6000, Sidney, British Columbia, Canada V8L 4B2.

Overstuffed

Jeremy J. D. Greenwood

Manual of Ornithology: Avian Structure and Function. By Noble S. Proctor and Patrick J. Lynch. Yale University Press: 1993. Pp. 340. £30, \$45.

PROCTOR and Lynch believe that there is a need for a laboratory manual "to collect under one cover the visual references necessary for an undergraduate laboratory course in ornithology". Given the limited number of illustrations in textbooks, there may well be such a need. If so, the illustrations in this book fill it. Not only do they portray avian morphology in detail but they do so clearly, making reference to specimens straightforward.

Unfortunately, the book contains no dissection instructions, although it describes how to prepare study skins, surely a much less important skill than basic dissection. One of the clearest instructions for preparing a study skin is: 'Do not overstuff the specimen'. It is advice that the authors would have done well to follow. For example, a chapter on field techniques fits uncomfortably in what is otherwise strictly a laboratory manual. Another element of overstuffing is that there is too much repetition in both the text and the illustrations — eight pictures to show general topography, for example, with three more for the details of the wing and four for the head. But the main way in which the book has been extended is in the background and supporting material. The authors admit that they found it difficult to walk the tightrope between textbook and reference manual. They have chosen to include much material extending beyond the basic morphology appropriate to a laboratory manual, in an attempt to relate the morphology to the life and natural history of birds. Unfortunately, this material is too brief and too idiosyncratic to be any substitute for an ornithology textbook, yet it is voluminous enough to get in the way of using the book as a laboratory manual. There may be a need for a laboratory manual in ornithology, but this does not meet it in a useful way. □

Jeremy J. D. Greenwood is in the British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU, UK.

Using small perturbations to control chaos

Troy Shinbrot, Celso Grebogi, Edward Ott & James A. Yorke

The extreme sensitivity of chaotic systems to tiny perturbations (the 'butterfly effect') can be used both to stabilize regular dynamic behaviours and to direct chaotic trajectories rapidly to a desired state. Incorporating chaos deliberately into practical systems therefore offers the possibility of achieving greater flexibility in their performance.

CHAOTIC systems are characterized by extreme sensitivity to tiny perturbations. This characteristic, known as the 'butterfly effect', is often regarded as a troublesome property, and for many years it has been generally believed that chaotic motions are neither predictable nor controllable. The first reference we find to a differing view is due to John von Neumann, who is reported¹ to have stated around 1950 that small, carefully chosen, pre-planned atmospheric disturbances could lead after some time to desired large-scale changes in the weather. Although this specific application might be problematic, the basic idea of using chaotic sensitivity seems to have been clearly appreciated by von Neumann. Here we review recent work which demonstrates that the butterfly effect permits the use of tiny feedback perturbations to control trajectories in chaotic systems—a capability without a counterpart in nonchaotic systems. These considerations apply to cases in which the chaotic dynamics can in principle be defined by only a few variables, so systems where there are many active degrees of freedom (for example, the weather, and high-Reynolds-number flows) may not be tractable. However, we emphasize that cases of high (or infinite) dimensional systems for which the 'attractor' (and hence the dynamics) is low dimensional are common.

The research that we will review fits broadly into two categories. First we will discuss how, as proposed in ref. 2, one can select a desired behaviour from among the infinite variety of behaviours naturally present in chaotic systems, and then stabilize this behaviour by applying only tiny changes to an accessible system parameter (related work appears in refs 3–32, 77, 78). Moreover, we will show how one can switch between behaviours as circumstances change, again using only tiny perturbations. This means that chaotic systems can achieve great flexibility in their ultimate performance. Second, we will show how one can use the sensitivity of chaotic systems to direct trajectories rapidly to a desired state^{33–38}. For example, a few years ago, NASA scientists used only small amounts of residual hydrazine fuel to send the spacecraft ISEE-3/ICE more than 50 million miles across the Solar System, thereby achieving the first scientific cometary encounter^{39–43}. This feat was made possible by the sensitivity of the three-body problem of celestial mechanics to small perturbations, and would not have been possible in a nonchaotic system, in which a large effect typically requires a large control^{44–46}.

Stabilizing unstable orbits

One of the fundamental aspects of chaos is that many different possible motions are simultaneously present in the system. A particular manifestation of this is the fact that there are typically an infinite number of unstable periodic orbits that co-exist with the chaotic motion^{47,48}. By a periodic orbit, we mean an orbit that repeats itself after some time (the period). If the system were precisely on an unstable periodic orbit, it would remain on that orbit forever. These orbits are unstable in the sense that the smallest deviation from the periodic orbit (for example due to noise) grows exponentially rapidly in time, and the system orbit quickly moves away from the periodic orbit. Thus, although these periodic orbits are present, they are not typically observed. Rather, what one sees is a chaotic trajectory which bounces

around in an erratic, seemingly random fashion. Very rarely, the chaotic trajectory may, by chance, closely approach a particular unstable periodic orbit, in which case the chaotic trajectory would approximately follow the periodic cycle for a few periods, but it would then rapidly move away because of the instability of the periodic orbit. In addition to periodic orbits, it is common for continuous time dynamical systems to have unstable steady states embedded in chaotic motion (see our subsequent discussion of the Lorenz attractor). A ball placed exactly at the top of a hill is an example of an unstable steady state.

Although the existence of steady states and an infinity of different unstable periodic orbits embedded in chaotic motion is not usually obvious in free-running chaotic evolution, these orbits offer a great potential advantage if one wants to control a chaotic system. To demonstrate this, we adopt the following strategy². First we examine the unstable steady states and low-period unstable periodic orbits embedded in the chaotic motion. For each of those unstable orbits, we ask whether the system performance would be improved if that orbit were actually followed. We then select one of the unstable orbits that yields improved performance. Assuming the motion of the free-running chaotic orbit to be ergodic, eventually the chaotic wandering of an orbit trajectory will bring it close to the chosen unstable periodic orbit or steady state. When this occurs, we can apply our small controlling perturbations to direct the orbit to the desired periodic motion or steady state. Moreover, if a small amount of noise is present, we can repeatedly apply the perturba-

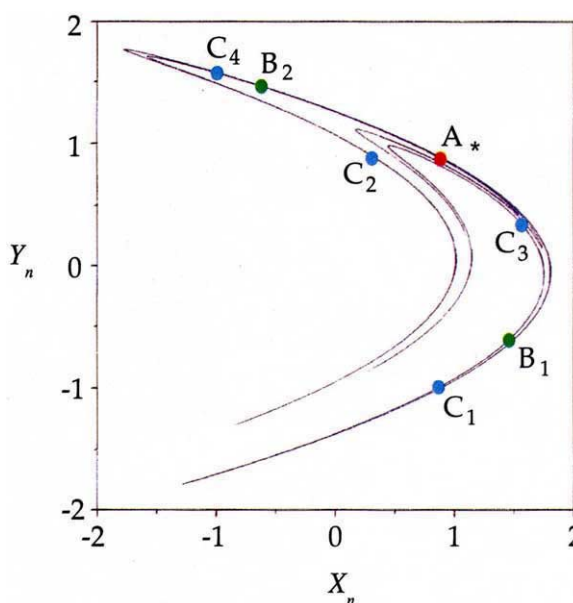


FIG. 1 Hénon attractor, with period-1 point, A_* ; period-2 points, B_1 and B_2 ; and period-4 points, C_1 , C_2 , C_3 and C_4 .

tions to keep the trajectory on the desired orbit. Thus small, carefully chosen perturbations are able to effect a large beneficial change in the long-term system behaviour. If, on the other hand, the system dynamics were not chaotic but were, say, stable periodic, then small controls could only change the orbit slightly. We would then be closely restricted to whatever system performance the stable periodic orbit gave, and we would have no option for improvement using small controls.

Furthermore, one may want a system to be used for different purposes or under different conditions at different times. If the system is chaotic, this requirement might be accommodated without altering the gross configuration of the system. In particular, depending on the use desired, the system could be optimized by switching the temporal programming of the small controls to stabilize different orbits. By contrast, in the absence of chaos, completely separate systems might be required for each use. Thus, when designing a system, it may be advantageous

to build chaos in, so as to achieve flexibility. For an experimental example, see Box 1. This experiment couples a magnetic field with a gravitationally buckling ribbon. Other recent relevant experiments have involved laser systems¹⁰⁻¹⁴, electrical circuits¹⁵, thermal convection¹⁶ and arrhythmically oscillating cardiac tissue (controlled using a small electrical stimulus)¹⁸.

We will now consider examples of dynamical processes that we wish to control. A simple example might be a metre stick balanced on one's palm. This system is not chaotic and, provided that the stick does not stray too far from the vertical, it can be stabilized in its normally unstable, vertical state by making small motions of one's palm. This is an example of stabilizing an unstable steady state.

As a simple illustrative example of controlling a chaotic system we consider the 'Hénon map'⁴⁹. Here the word 'map' refers to the fact that the time variable is discrete and integer-valued. The Hénon map is also described as a 'two-dimensional map'

BOX 1 Experimental confirmation

THE control of chaos by the application of tiny perturbations has been experimentally applied in several laboratories^{14,17,29} (a modification, effective in stabilizing high-period orbits, appears in refs 10, 15). The first such application⁹ used a nonlinear, inverted, magneto-elastic ribbon. This ribbon, sketched in Fig. B1, was clamped at its base but was otherwise free to move. The stiffness of the material was nonlinearly dependent on applied magnetic field, and in an oscillating field, applied by external field coils (not shown), the ribbon alternately buckled and stiffened in a complicated, chaotic pattern⁷⁶. The position of the ribbon, $X(t)$, was measured at a point near its base using an optical sensor. By making small adjustments to the amplitude of the oscillating field (variations were less than 9% of its nominal value), the ribbon could be trapped in a variety of periodic motions. Figure B2 shows the ribbon displacement sampled once every oscillation of the applied magnetic field as a function of the number of periods, n . The sampling of the orbit at the drive period can be regarded as a "stroboscopic" surface of

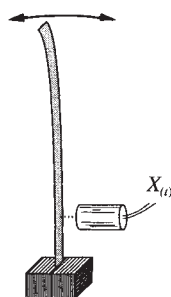


FIG. B1 Chaotic ribbon.

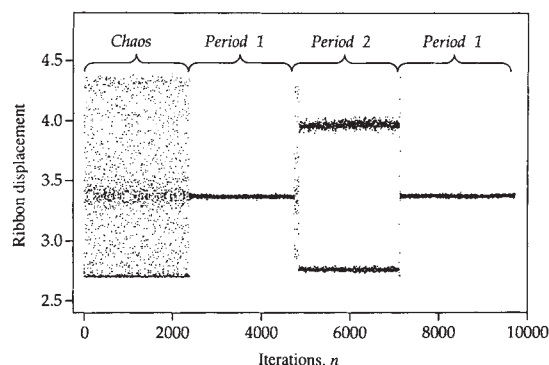


FIG. B2 Experimental stabilization of chaotic ribbon (reprinted with permission from ref. 9).

section. Initially no control was applied, leading to chaotic variation in the range of ribbon displacements between 2.7 and 4.4 on the vertical axis. Then, at about time $n \approx 2,200$, the control algorithm was activated to attempt to stabilize a period-1 orbit in the stroboscopic surface of section (compare this with Fig. 2b). After this orbit had been stabilized for over 2,000 oscillations, the control algorithm was switched at time $n \approx 4,900$ to try to stabilize a period-2 orbit (see Fig. 2c), and was later switched back to the period-1 orbit at $n \approx 7,100$.

The nonlinear ribbon was subsequently used to show that chaotic sensitivity can be used to actually direct trajectories in a chaotic system. By making small adjustments to the applied magnetic field (less than 5% of its nominal value), any trajectory could be quickly directed to a small target state. Figure B3 shows several successive trajectories which are rapidly brought from a variety of initial states, indicated by grey circles, to the target neighbourhood, $X = 2.5 \pm 0.01$. This target was chosen because it is ordinarily seldom visited: without control, the time to reach the chosen target was once in 500 iterations on average. By applying small perturbations, on the other hand, the target could typically be reached in less than once every 20 iterations.

We reiterate that the control was accomplished in real time and in the presence of experimental noise and modelling errors (to accommodate this, the targeting algorithm must be periodically reapplied; see refs 33, 35 for details). Additionally, the computational models used both for stabilization and for targeting were constructed entirely from available experimental data and without an *a priori* analytic model.

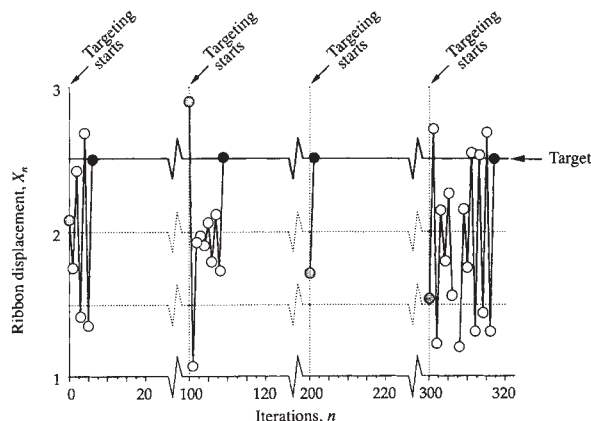


FIG. B3 Several successive realizations of targeting of a small neighbourhood of $X = 2.5$. Ribbon is allowed to wander chaotically between targeting attempts; targeting is initiated at $n = 0, 100, 200, 300$. With targeting, the neighbourhood of $X = 2.5$ is reached within 20 iterations; without targeting, the same neighbourhood is visited less than once every 500 iterations. The origin of the ordinate scale here differs from that of Fig. B2 because the optical sensor was moved between the times of the two experiments.

because the state of the system at time n (where $n = 0, 1, 2, \dots$) is given by two scalar variables, x_n and y_n . The map specifies a rule for evolving the state of the system at time n to the state at time $(n+1)$. For the Hénon map, this rule is

$$x_{n+1} = p + 0.3y_n - x_n^2 \quad (1a)$$

$$y_{n+1} = x_n \quad (1b)$$

where the parameter p is set to a nominal value of $p_0 = 1.4$. Thus, given an initial state (x_1, y_1) , the map allows us to calculate (x_2, y_2) , which, when again inserted into the map, yields (x_3, y_3) , and so on. Iterations of this map eventually converge to a strange attractor, as shown in Fig. 1. This structure is called an attractor because distant points are drawn toward it under successive iterations of the map, and it is called strange because it is infinitely intricate on every distance scale we might choose to examine. In the terminology of Mandelbrot, it is a fractal and has a fractal dimension of ~ 1.3 . In the figure, we show just a few of the infinite number of embedded unstable periodic orbits; a period-1 point, A_* , which is revisited every map iteration, period-2 points, B_1 and B_2 , each of which are revisited every other map iteration ($B_1 \rightarrow B_2 \rightarrow B_1 \rightarrow B_2 \dots$), and period-4 points, $C_1 \dots C_4$, which are cycled through every four map iterations.

Although the above example is a map (the time, n , is discrete), many problems in science and engineering involve continuous time, such as a system of M first-order, autonomous ordinary differential equations, $d\xi/dt = G(\xi)$, where $\xi(t) = (\xi^{(1)}(t), \xi^{(2)}(t), \xi^{(3)}(t), \dots, \xi^{(M)}(t))$ is an M -vector, and the continuous variable t denotes time. In such a case, discrete time systems are still of interest, as the M -dimensional continuous time system can be reduced to an $M-1$ dimensional map by the Poincaré surface-of-section technique illustrated in Fig. 2a for $M=3$. Here we associate the continuous time trajectory with a discrete time trajectory, $Z_1, Z_2, \dots$, where Z_n denotes an $M-1$ coordinate vector specifying the position on the surface of section of the n th upward piercing of the surface. Given a Z_n , we can integrate the equation $d\xi/dt = G(\xi)$ forward in time from that point, until the next upward piercing of the surface of section, at the surface coordinates Z_{n+1} . Thus Z_{n+1} is uniquely determined by Z_n , and there must exist a map, $Z_{n+1} = F(Z_n)$, from one trajectory point on the surface of section to the next. Although we may not be able to write down F explicitly, the knowledge that it exists is still useful. Figure 2b shows a periodic orbit of the continuous time system which results in a period-1 orbit of the associated Poincaré surface of section map: $Z_* = F(Z_*)$. Figure 2c shows a periodic orbit of the continuous time system resulting in a period-2 orbit of the map: $Z_2 = F(Z_1)$, $Z_1 = F(Z_2)$.

Let us say that we have selected one of these unstable periodic orbits as providing the best performance of some hypothetical system. For simplicity, we consider the case where the desired orbit is a period-1 orbit of some N -dimensional map

$$Z_{n+1} = F(Z_n, p) \quad (2)$$

where Z is an N -dimensional vector, and p denotes a system parameter. We first approximate the dynamics near the period-1 point, denoted Z_* (where $Z_* = F(Z_*, p_0)$), for values of the parameter, p , close to the nominal value, p_0 , by the linear map

$$(Z_{n+1} - Z_*) = A \cdot (Z_n - Z_*) + B \cdot (p - p_0) \quad (3)$$

Here A is an $N \times N$ dimensional jacobian matrix and B is an N -dimensional column vector, where $A = \partial F / \partial Z$, $B = \partial F / \partial p$, and these partial derivatives are evaluated at $Z = Z_*$ and $p = p_0$. We now assume that we can adjust the parameter p on each iteration. That is, we determine Z_n and on that basis make an appropriate small change in p from the nominal value p_0 . Thus we replace p by p_n . Taking the control law to be linear, we have

$$(p_n - p_0) = -K^T \cdot (Z_n - Z_*) \quad (4)$$

Here K is a constant N -dimensional column vector and K^T is

its transpose. Choice of the vector K specifies the control law specifying p_n on each iteration. Substituting equation (4) into equation (3), we obtain

$$\delta Z_{n+1} = (A - B \cdot K^T) \delta Z_n \quad (5)$$

where $\delta Z_n = Z_n - Z_*$. Thus the period-1 point, Z_* , will be stable if one can choose K so that the matrix $(A - B \cdot K^T)$ only has eigenvalues with modulus smaller than unity. In this case, $\delta Z_n \rightarrow 0$ (that is, $Z_n \rightarrow Z_*$) as $n \rightarrow \infty$. The choice of K represents a standard problem in control theory (for example, see ref. 50). (In fact, the matrix $(A - B \cdot K^T)$ can be made to have any desired set of N eigenvalues provided that B satisfies a certain condition (called controllability⁵⁰.) This procedure gives the time-dependent parameter values, p_n , required to stabilize an unstable period-1 point in a chaotic system. For the case of higher-period orbits, see ref. 4. We imagine that, because of practical constraints, we cannot make the deviations of p_n from p_0 too large. Thus we imagine that $|p_n - p_0|$ is bounded by some allowed maximum value, $\delta p_{\max}$. Hence by equation (4), $\delta p_{\max} > |K^T \cdot (Z_n - Z_*)|$. If the system state is outside this region, we apply no perturbation and wait until the state falls within the

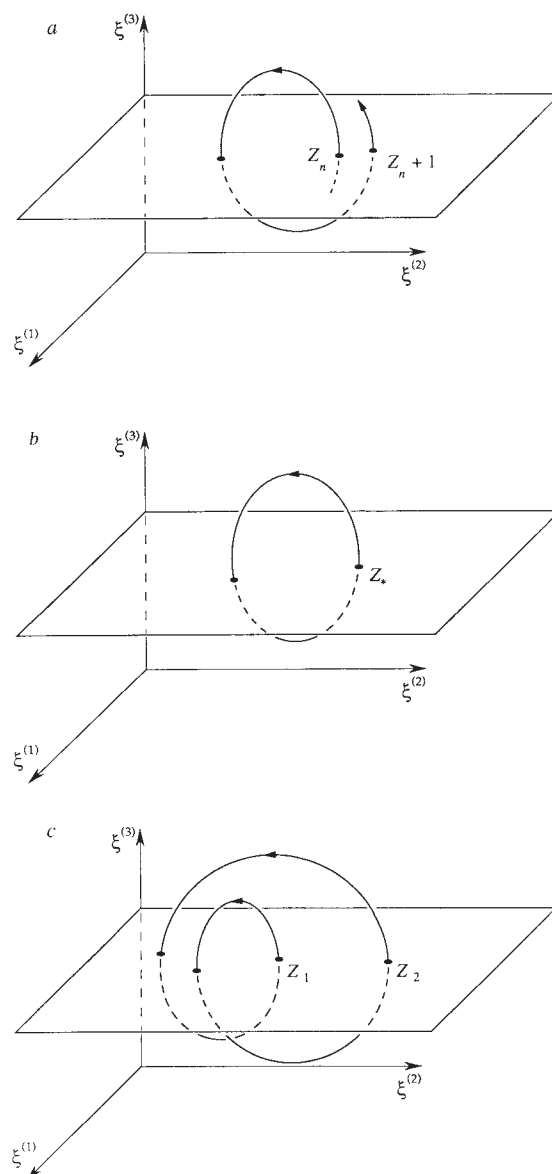


FIG. 2 a, The Poincaré surface-of-section technique. The surface of section, here shown for the case $\xi_3 = \text{constant}$, can, in principle, be chosen in any convenient way. b, Period-1 orbit; c, period-2 orbit.

given region. Once this occurs, we trap the system near the desired periodic state by applying the appropriate small nudges given by equation (4). Stabilization of the periodic points of the Hénon map by small perturbations is similar to stabilizing a vertical metre stick insofar as neither can be accomplished if the current state is far from the desired unstable state. We will discuss later how to direct a chaotic orbit to a neighbourhood of a desired state.

To illustrate stabilization of an unstable period-1 orbit in a more intuitive and geometrical manner, we consider the special case of a two-dimensional map where the desired period-1 orbit has one unstable and one stable direction. That is, there are two curves through Z_* , called the stable manifold and the unstable manifold, as shown in Fig. 3a. We consider a small neighbourhood of Z_* , so that the lines shown in Fig. 3 are roughly straight. An orbit starting from a point on the stable manifold remains on the stable manifold and moves exponentially toward Z_* , whereas orbits of points on the unstable manifold move exponentially away from Z_* . This corresponds to the case in which the matrix A has two real eigenvalues, one with magnitude less than one (stable), and one with magnitude greater than one (unstable). Also shown in Fig. 3a is a dashed line along which Z_* can be shifted by a change of the parameter p (this line is parallel to the vector B).

Imagine that the point Z_n falls close to $Z_*(p_0)$, as shown in Fig. 3a. We now perturb the value of p from p_0 to $p_0 + \delta p$, as shown in Fig. 3b. On the next iteration, the orbit is attracted towards $Z_*(p_0 + \delta p)$, in a direction parallel to its stable manifold, and, at the same time, is repelled from $Z_*(p_0 + \delta p)$ in a direction parallel to its unstable manifold. Thus if δp is chosen properly, we can cause Z_{n+1} to fall precisely on the stable manifold of $Z_*(p_0)$. Thereafter, we can return the parameter to its nominal value, p_0 , and the orbit will remain on the stable manifold of $Z_*(p_0)$ and will approach $Z_*(p_0)$. In terms of the more general discussion leading to equations (5), the above corresponds to a special choice of the control vector, K , such that one of the eigenvalues of $(A - B \cdot K^T)$ is zero and the other is the original stable eigenvalue of A . As shown in ref. 4, this choice is optimal in that it minimizes the average time during which the orbit wanders chaotically before it can be stabilized.

As an example, in Fig. 4, we show the results of stabilizing the periodic orbit, A_* , on the Hénon attractor by adjusting p by less than 1% of its nominal value. Starting at a random initial point on the attractor, we see that for the first 86 iterations, the trajectory moves chaotically on the attractor, never falling within the desired small region about A_* . Then, on the 87th iteration, the state falls within the desired region, and thereafter is held

near A_* . In the presence of noise, the stabilization procedure would have to be regularly reapplied (see ref. 3 for details).

The control strategy need not rely on an *a priori* analytical model. Just as one can balance a metre stick on the palm without knowing anything at all about Newton's equations of motion for the stick, one can produce effective stabilization for more complicated systems without an explicit set of differential equations. By empirical study of the effects of small parameter changes on orbits near a desired periodic state, an arbitrary periodic state in a chaotic attractor can be stabilized using only small controls (see Box 1). This is important because in experimental situations one may often not have on hand an accurate analytical model.

Figure 5 illustrates how knowledge of a periodic orbit, the matrix A , and the vector B can all be extracted purely from observations of the trajectory on the strange attractor. Imagine that we collect a long data string of observed surface-of-section piercings, Z_1, Z_2 and so on. If two successive Z s are close to each other, say Z_{100} and Z_{101} , then there will typically be a period-1 orbit Z_* nearby^{47,48} (Fig. 5). Having observed a first such close return, we then search the succeeding data for other close-return pairs (Z_n, Z_{n+1}) restricted to the small region of the original close return (shown shaded in Fig. 5). Because orbits on a strange attractor are ergodic, we will have many such pairs if the data string is long enough. Assuming the shaded region to be small, we then try to fit these close returning pairs with a linear relation

$$Z_{n+1} = \hat{A} \cdot Z_n + \hat{C} \quad (6)$$

Generally, if there is noise in the data, we would want to use as many pairs as possible and fit the matrix $\hat{A}$ and the vector $\hat{C}$ using a least-squares fit to the data. Thus $\hat{A}$, the least-squares fit matrix, is an approximation to the jacobian matrix A of equation (3), and the period-1 point, $Z_*(p)$, is approximated by $(1 - \hat{A})^{-1} \cdot \hat{C}$. To find B of equation (3), we change p slightly, $p \rightarrow p + \Delta p$, redetermine the period-1 point $Z_*(p + \Delta p)$ as before, and approximate B as $[Z_*(p + \Delta p) - Z_*(p)] / \Delta p$. To find period-2 orbits, we proceed in the same way, but for pairs (Z_n, Z_{n+2}) that are close after two surface-of-section piercings (and similarly for higher periods).

In experimental studies of chaotic dynamics, it is often helpful to use 'delay coordinate embedding'^{51,52}. These allow one to obtain information on the topological phase space structure of an attractor even when one cannot directly measure the instantaneous system state vector, ξ , which we take to be M -dimensional. As a typical example, the time history of only a single scalar variable, say $\phi(t)$, might be the only thing that one

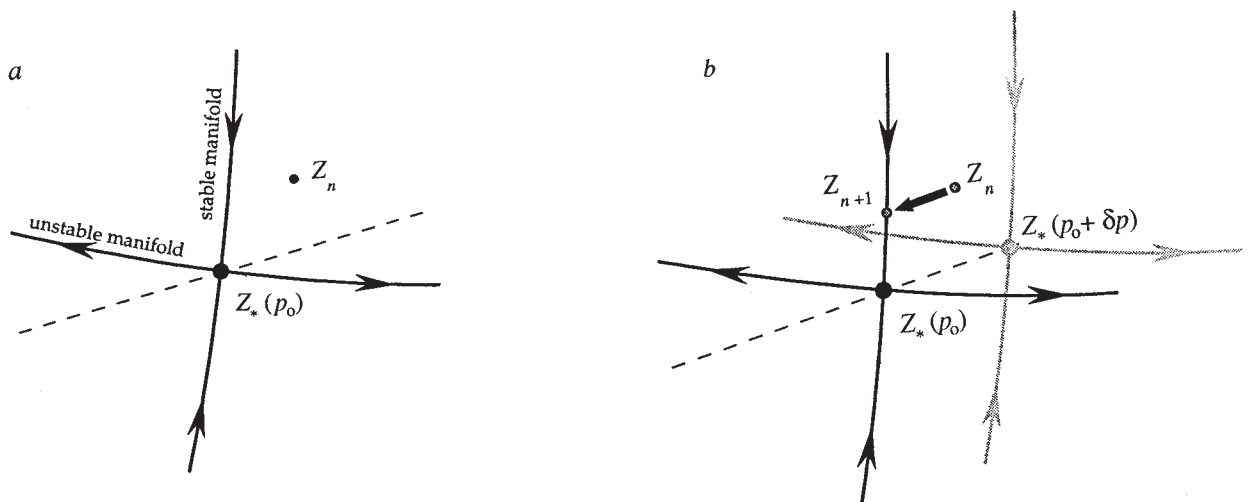


FIG. 3 a, The period-1 point, $Z_*(p_0)$, its stable and unstable manifolds (solid lines), and the line (dashed) along which Z_* can be shifted by perturbation of the parameter p . b, Result of perturbing p to $p_0 + \delta p$; the stable and

unstable manifolds of $Z_*(p_0 + \delta p)$ are shown as grey lines through $Z_*(p_0 + \delta p)$.

can measure experimentally. To proceed, one forms the Q -dimensional 'delay coordinate vector'

$$\Phi(t) = (\phi(t), \phi(t - T_D), \phi(t - 2T_D), \dots, \phi(t - (Q-1)T_D)) \quad (7)$$

where T_D is some conveniently chosen delay time. Embedding theorems^{53,54} guarantee that for $Q \geq 2M + 1$, the vector $\Phi(t)$ is generically a global one-to-one representation of the system state. (Actually, for our purposes we do not require a global embedding; we only require a one-to-one correspondence in the small region near the periodic orbit, and this can typically be achieved with $Q = M$.) To obtain a discrete time series from $\Phi(t)$, we can again use a Poincaré surface of section, but this time in Φ -space. Let $\tilde{Z}_n$ denote points in the Φ -space surface of section, and let the corresponding surface-of-section map for a constant value of the parameter p be denoted $\tilde{Z}_{n+1} = \tilde{F}(\tilde{Z}_n, p)$. As pointed out in ref. 7, in the presence of parameter variation ($p \rightarrow p_n$), delay coordinates lead to a map of a form other than $\tilde{Z}_{n+1} = \tilde{F}(\tilde{Z}_n, p_n)$, which is the form assumed for equations (3) and (4). As an example, say that the time interval T_n between the surface-of-section piercing at $\tilde{Z}_n$ and at $\tilde{Z}_{n+1}$ is such that

$$rT_n \geq (Q+1)T_D > (r-1)T_n \quad (8)$$

where r is an integer. Then the relevant map is of the form

$$\tilde{Z}_{n+1} = \tilde{F}(\tilde{Z}_n, p_n, p_{n-1}, \dots, p_{n-r}). \quad (9)$$

This follows because $\Phi(t_n)$, where t_n denotes the n th piercing of the surface of section (corresponding to $\tilde{Z}_n$), has components $\phi(t_n), \dots, \phi(t_n - (Q-1)T_D)$ and hence $\tilde{Z}_n$ must depend, not only on p_n , but also on all the other parameter values in effect during the time interval $t_n \leq t \leq [t_n - (Q-1)T_D]$. For discussion of how the analysis of equations (3) to (5) can be extended to the case of delay coordinates, see refs 4 and 7.

Although the approach just described provides a systematic means of choosing a feedback algorithm, it will often be the case (particularly when the dynamics are low-dimensional) that a trial and error procedure will succeed^{10,15}: simply choose some feedback law (arbitrarily choose K in equation (4)) and vary it until it is observed that the desired orbit stabilizes. (Knowledge of the periodic orbit location in phase space is still required, so a procedure such as that of Fig. 5 may still be necessary.)

Further discussion of stabilization

We now briefly discuss some other considerations relevant to the stabilization of unstable periodic orbits and steady states embedded in strange attractors. One issue is related to bifurcations. The word bifurcation, when applied to a periodic orbit or steady state, refers to a change in the character of the orbit from stable to unstable while the system is continuously changed. Often the onset of chaos comes about as a result of bifurcations of periodic orbits or steady states. Two well known cases are the commonly observed cascade of period-doubling bifurcations preceding chaos, and the onset of chaos when a steady state bifurcates from stable to unstable. Thus one way of controlling chaos is to prevent or change the character of these bifurcations. This has been discussed in refs 18 and 19, where control methods that are insensitive to errors in the knowledge of the system are used. In this regard, we also mention the laser experiments of refs 12 and 13.

Control can also be used as a means of tracking the location of unstable orbits as the system is changed^{21,22}. In a recent experiment, this technique was used to increase the stable steady power output of a laser by an order of magnitude¹⁷. Another issue is that of prescribing controls that are assured of bringing the trajectory to the desired periodic orbit or steady state. This has been addressed using the Lyapunov function method^{16,23}.

Other interesting work in this general area includes the use of the describing function technique of control theory for finding and controlling unstable periodic orbits²⁴; control of unstable chaotic sets (as opposed to chaotic attractors)²⁵; the control of homoclinic orbits²⁶; and control of aperiodic orbits⁵⁵.

Directing chaotic trajectories

Stabilizing a system by small perturbations is extremely effective once the system at hand comes close to the desired state. But if it starts far from the desired state, it might take an unacceptably long time before a typical orbit comes close enough to the desired state to be captured. We now discuss how small perturbations, applied when the orbit is far from the desired state, can be used to steer the system to this state. For example, chaotic rhythms in cardiac tissue can be stabilized as outlined above¹⁷. What if one wants to stabilize a regular rhythm in the heart without having the luxury of simply waiting for the system to fall near a desired state? It is, moreover, intrinsically desirable to be able to steer a system to a general target in phase space (not necessarily a periodic orbit). NASA's mission specialists demonstrated this when, as mentioned earlier, they achieved the first scientific cometary encounter by steering the spacecraft ISEE-3/ICE in a complex trajectory using only small nudges from the spacecraft's dwindling fuel supply³⁹⁻⁴³.

The application of chaotic sensitivity to steer trajectories to targets in chaotic systems using small controls is discussed in refs 33-37; related work appears in refs 38 and 55. To understand the idea in its simplest form, consider the well known logistic map (see for example ref. 56):

$$X_{n+1} = pX_n(1 - X_n) \quad (10)$$

where we take a nominal value of 3.9 for p . This map is used to describe the behaviour of a population of organisms after successive years, where n denotes the year. In this case, X represents the (normalized) population, and p defines its growth rate per year when the population is small. For any X between 0 and 1, it is easy, using only a pocket calculator, to adjust the growth rate slightly so that any given target, again between 0 and 1, is quickly reached.

Suppose for example that the current state is $X_1 = 0.4$ and we want to reach the vicinity of $X_n = 0.8$. If we can adjust p during the first year by a small amount, say between 3.8 and 4.0, then after one year, the population can range between $X_2 = 0.91$ and $X_2 = 0.96$. We then return to the nominal parameter value, $p = 3.9$, and after a second year the range grows to cover from $X_3 = 0.15$ to $X_3 = 0.31$. After a third year, the range grows even more to extend from $X_4 = 0.50$ to $X_4 = 0.84$. As our target, $X = 0.8$, is in this range, there must be some value of p_1 in the

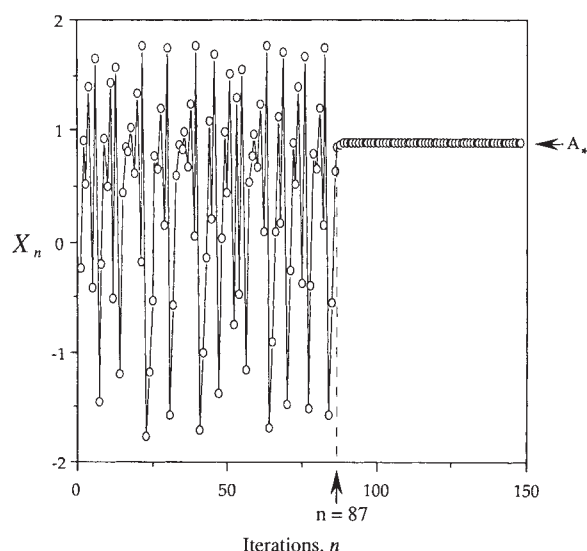


FIG. 4 Stabilization of the period-1 state for the Hénon map.

range $3.8 \leq p_1 \leq 4.0$ such that when p is shifted to that value in the first year, X falls on the target in only three years. Indeed, a little work with our pocket calculator reveals that we can reach our target by setting $p_1 = 3.83189\dots$. We emphasize that it is possible to accomplish this only because the logistic system is chaotic, and because chaotic systems are characterized by the exponential growth of small disturbances. This exponential growth implies that we can reach any accessible target extremely quickly (that is, in a time of the order of the logarithm of the maximum allowed size of the small parameter perturbation), using only a small perturbation.

As another example, consider the equations

$$\frac{dX}{dt} = \sigma(Y - X) \quad (11a)$$

$$\frac{dY}{dt} = -XY - Y + r + p(t) \quad (11b)$$

$$\frac{dZ}{dt} = XY - bZ \quad (11c)$$

which (for $p(t) = 0$) were introduced⁵⁷ by Lorenz as a simplified model of chaotic fluid thermal convection. The Lorenz equations provide a leading-order description of the dynamics of a fluid contained in a thin vertically oriented torus with a heat source applied at the bottom⁵⁸. The equations with $p(t) = 0$ have a strange attractor for the parameter values $\sigma = 10$, $r = 28$, $b = 8/3$, and an orbit on this chaotic attractor is shown in blue in Fig. 6. The steady state, $X(t) = Y(t) = Z(t) = 0$, representing no fluid convection, is a solution of the Lorenz equations. Furthermore, this solution is contained in the chaotic attractor. Note from Fig. 6 that the blue finite-duration orbit does not reach anywhere near this steady state (the open circle in the figure). If the orbit were followed long enough, it would eventually come arbitrarily near the stationary state. We estimate that of the order of one in 10^{10} orbits around either of the lobes of the attractor ever passes through a sphere of radius 0.1 centred at $X = Y = Z = 0$. (For comparison, the blue trajectory in Fig. 6 shows only about 20 orbits around either lobe of the attractor.) For the purpose of demonstrating control of the Lorenz system we have added the term $p(t)$ to the right-hand side of equation (11b). For the physical situation of a vertically oriented fluid-filled torus, the term $p(t)$ represents a perturbation of the position of the heat source slightly to the left or right in the plane of the torus ($p = 0$ corresponds to heating exactly at the bottom of the torus). Restricting the perturbations, $p(t)$, to be small, $|p(t)| < 0.01$, it typically takes of the order of 10 orbits around a lobe to reach the origin (as compared, again, with 10^{10} orbits to reach $\sqrt{X^2 + Y^2 + Z^2} < 0.1$ when no control is applied). Figure 6 shows such an orbit in red. The method for programming the perturbations is described in ref. 36 and is similar in principle to that which we have described for the logistic map, equation (10).

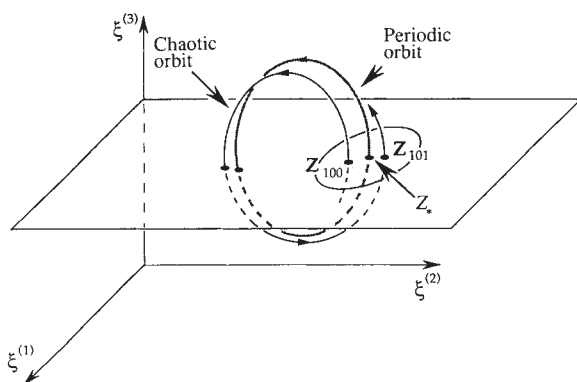


FIG. 5 Determining unstable periodic point, $\mathbf{Z}$ from data, $\mathbf{Z}_{100}, \mathbf{Z}_{101}, \dots$

The preceding example illustrates the idea of targeting. More detailed descriptions and methods can be found in refs 33–38. One technique is to patch together different trajectories to reach a desired target state. These trajectories can originate in a single chaotic system with differing parameter values^{37,38}, or from a systematic examination of small changes in the system state^{39–43}. NASA's manoeuvre of the spacecraft ISEE-3/ICE is such a case. Five separate swings past the Moon were combined to produce the final spacecraft trajectory. By inventive choices of intermediate trajectories, researchers have shown this technique to be effective in matching greatly differing initial and final state vectors.

At this point, we must make two remarks. First, the sensitivity of chaotic systems allows us to produce large changes in the orbit of the system after some time using tiny perturbations, but the same sensitivity makes the final state depend on ubiquitous noise. The noise, if small enough, can be compensated for, however, by periodically reapplying the targeting algorithm, thereby obtaining mid-course corrections of the parameter perturbation (see Box 1 and refs 33 and 35). Second, one needs a global model of the system in order to direct trajectories. This differs from the stabilization of periodic orbits or steady states, where one only needs local information, near the desired periodic orbit. In general, this makes the use of experimental delay coordinate techniques for application to targeting more difficult. Nevertheless, in some cases it may be possible (Box 1).

The control of chaos by small perturbations has been developed into a proposed application for communication⁵⁹. Consider the 'double scroll' electrical oscillator⁶⁰, which yields a chaotic signal consisting of a seemingly random sequence of positive and negative peaks. If we associate a positive peak with a one and a negative peak with a zero, the signal yields a binary sequence. With small control perturbations, we can cause the signal to follow an orbit whose binary sequence represents the information we wish to communicate. (Here our 'target' is a particular binary sequence rather than a point in phase space.) Hence the chaotic power stage that generates the waveform for transmission can remain simple and efficient (complex chaotic behaviour occurs in simple systems), while all the complex electronics controlling the output can remain at the low-power microelectronic level.

Other ways to alter chaotic dynamics

We have focused here on work using the sensitivity of chaotic systems to stabilize existing chosen periodic orbits and steady states and to steer trajectories with only small controls. Several authors have developed other techniques to alter the dynamics of chaotic systems without explicitly using this sensitivity, and in this section, we summarize some of their key contributions.

One body of research^{61–68} seeks to control a nonlinear system to follow a prescribed goal dynamics. If we denote the system by

$$\frac{d\mathbf{g}}{dt} = \mathbf{F}(\mathbf{g}) + \mathbf{U}(t) \quad (12)$$

where $\mathbf{U}(t)$ is an additive controlling term, then the object is to choose $\mathbf{U}(t)$ so that $|\mathbf{g}(t) - \mathbf{g}| \rightarrow 0$ as $t \rightarrow \infty$, where $\mathbf{g}(t)$ is the goal dynamics. To accomplish this, the simple choice

$$\mathbf{U}(t) = \frac{d\mathbf{g}}{dt} - \mathbf{F}(\mathbf{g}(t)) \quad (13)$$

is made. Thus $\mathbf{g}(t) = \mathbf{g}$ is clearly a solution of the controlled equations. What is not so clear is that convergence to this goal will often occur ($|\mathbf{g}(t) - \mathbf{g}| \rightarrow 0$ as $t \rightarrow \infty$). Whether it does so depends on the particular $\mathbf{F}$ and the initial condition, $\mathbf{g}(0)$. The regions of $\mathbf{g}$ -space such that controlled orbits originating in them converge to the goal, $\mathbf{g}$, are called entrainment regions^{64–66}. This method potentially works for nonlinear systems in general (not necessarily chaotic) and has the advantage of not requiring

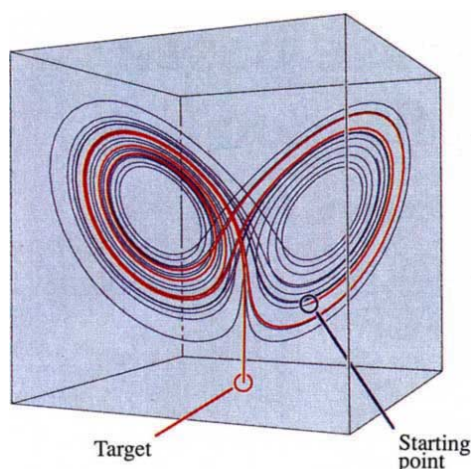


FIG. 6 Targeting of steady state in Lorenz attractor. Blue; trajectory without control; red: trajectory with control.

feedback. On the other hand, the applied controls are not typically small and convergence to the goal is not assured.

Another body of research addresses the effects of periodic⁶⁹⁻⁷³ and stochastic^{74,75} perturbations on chaotic systems. As one might expect, the effects of such perturbations can be quite difficult to predict in general, and indeed these studies are not 'goal oriented', in that a desired behaviour is not specified in advance and a generic technique for achieving such a goal has not been developed. Nevertheless, dramatic changes in the dynamics of chaotic systems have been recorded using these methods; for example, periodic or nearly periodic behaviour can sometimes be produced from originally chaotic dynamical systems.

The advantages of chaos

The presence of chaos may be a great advantage for control in a variety of situations. In a nonchaotic system, small controls typically can only change the system dynamics slightly. Short of applying large controls or greatly modifying the system, we are stuck with whatever system performance already exists. In a chaotic system, on the other hand, we are free to choose between a rich variety of dynamical behaviours. Thus we anticipate that it may be advantageous to design chaos into systems, allowing such variety without requiring large controls or the design of separate systems for each desired behaviour.

The general problem of controlling chaotic systems is very rich, and may help solve technologically important problems in widely diverse fields of study. In communications, it has been proposed that chaotic fluctuations can be put to use to send controlled, pre-planned signals⁵⁹. In physiology, applications have been proposed for controlling chaos in the heart¹⁷ and in neural information processing²⁷. In fluid mechanics, it has been demonstrated in a simple configuration that chaotic convection¹⁶ can be controlled. Chemical researchers have developed mechanisms for controlling chaotic autocatalytic reactions²⁸. Chaotic lasers^{10,11} have been controlled, as has the chaotic diode circuit¹⁵. The wealth of results such as these encourage us to look forward to a fruitful future for the control of chaotic systems. □

Troy Shinbrot is at the Department of Physics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA; present address for Troy Shinbrot is Laboratory for Fluid Mechanics, Chaos and Mixing, Northwestern University, Evanston, Illinois 60208, USA. Celso Grebogi and James A. Yorke are at the Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA. Edward Ott is at the Department of Electrical Engineering, the Department of Physics and the Institute for Systems Research, University of Maryland, College Park, Maryland 20742, USA.

- Dyson, F. *Infinite in All Directions*, 183-184 (Harper and Row, New York, 1988).
- Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **64**, 1196-1199 (1990).
- Ott, E., Grebogi, C. & Yorke, J. A. in *Chaos: Soviet-American Perspectives on Nonlinear Science* (ed. Campbell, D. K.) 153-172 (Am. Inst. Phys., New York, 1990).
- Ott, E., Romeiras, F. J., Grebogi, C. & Dayawansa, W. P. *Physica D* **58**, 165-192 (1992).
- Auerbach, D., Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Rev. Lett.* **69**, 3479-3482 (1992).
- Fowler, T. B. *IEEE Trans. Auto. Control* **34**, 201-205 (1989).
- Dressler, U. & Nitsche, G. *Phys. Rev. Lett.* **68**, 1-4 (1992).
- Pyragas, K. *Phys. Lett. A* **170**, 421-428 (1992).
- Ditto, W. L., Raue, S. N. & Spano, M. L. *Phys. Rev. Lett.* **65**, 3211-3214 (1990).
- Roy, R., Murphy, T. W., Maier, T. D., Gillis, A. & Hunt, E. R. *Phys. Rev. Lett.* **68**, 1259-1262 (1992).
- Gillis, Z., Iwata, C., Roy, R., Schwartz, I. B. & Triandaf, I. *Phys. Rev. Lett.* **69**, 3169-3172 (1992).
- Bielawski, S., Bouazaoui, M., Derozier, D. & Glorieux, P. *Proc. Nonlinear Dynam. Optical Syst. Topical Mtg. Alpbach, Austria, June 22-26* (Optical Society of America, Washington, DC 1992).
- Bielawski, S., Bouazaoui, M., Derozier, D. & Glorieux, P. *Stabilization and Characterization of Unstable Steady States in a Laser* (preprint, 1992).
- Reyl, C., Flepp, L., Badii, R. & Brun, E. *Control of NMR-laser Chaos in High-dimensional Embedding Space* (preprint, 1992).
- Hunt, E. R. *Phys. Rev. Lett.* **67**, 1953-1955 (1991).
- Singer, J., Wang, Y. Z. & Bau, H. H. *Phys. Rev. Lett.* **66**, 1123-1125 (1991).
- Garfinkel, A., Spano, M. L., Ditto, W. L. & Weiss, J. N. *Science* **257**, 1230-1235 (1992).
- Abed, E. H., Wang, H. O. & Lee, H.-C. *Proc. 1992 Am. Control Conf.*, 2236-2237 (Chicago, 1992).
- Wang, H. O. & Abed, E. H. *Proc. 2nd IFAC Workshop Syst. Struct. Control*, 494-497 (Prague, 1992).
- Wang, H. O. & Abed, E. H. *Proc. 2nd IFAC Nonlinear Control Sys. Design Symp.*, 57-62 (Bordeaux, France, June 1992).
- Schwartz, I. B. & Triandaf, I. *Phys. Rev. A* **46**, 7439-7444 (1992).
- Carroll, T., Triandaf, I., Schwartz, I. B. & Pecora, L. *Phys. Rev. A* **46**, 6189-6192 (1992).
- Chen, G. & Dong, X. *J. Circ. Syst. Comput.* **3** (in the press).
- Genesio, R. & Tesi, A. *Automatica* **28**, 531-548 (1992).
- Tél, T. *J. Phys. A* **24**, L1359-L1368 (1992).
- Bloch, A. M. & Marsden, J. E. *Theor. comput. Fluid Dynam.* **1**, 179-190 (1989).
- Ding, M. & Kelso, J. A. S. in *Measuring Chaos in the Human Brain* (eds Duke, D. & Pritchard, W.) 17-31 (World Scientific, Singapore, 1991).
- Petrov, V., Gáspár, V., Masere, J. & Showalter, K. *Nature* **361**, 240-243 (1993).
- Hübler, B., Doerner, R. & Martienssen, W. *Controlling Chaotic Motion in Noisy Systems* (preprint, Phys. Inst. der J. W. Goethe-Universität, Frankfurt-am-Main, 1992).
- Vassiliadis, D. *Physica D* (in the press).
- Huberman, B. A. & Lumer, E. *IEEE Trans. Circ. Syst.* **37**, 547-550 (1990).
- Sinha, S. & Ramaswamy, R. *Physica D* **43**, 118-128 (1990).
- Shinbrot, T., Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **65**, 3215-3218 (1990).
- Shinbrot, T., Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. A* **45**, 4165-4168 (1992).
- Shinbrot, T., Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **68**, 2863-2866 (1992).
- Shinbrot, T., Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Lett. A* **169**, 349-354 (1992).
- Kostelich, E., Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Rev. E* **47**, 305-310 (1993).
- Bradley, E. in *Lecture Notes Control and Information Sciences*, No. 165 (eds Jacob, G. & Lamnabhi-Lagarrique, F.) 307-325 (Springer, Berlin, 1991).
- Farquhar, R., Muhonen, D. & Church, L. C. *J. astronaut. Sci.* **33**, 235-254 (1985).
- Muhonen, D., Davis, S. & Dunham, D. *J. astronaut. Sci.* **33**, 255-273 (1985).
- Dunham, D. W. & Davis, S. A. *J. astronaut. Sci.* **33**, 275-288 (1985).
- Muhonen, D. & Foltz, D. *J. astronaut. Sci.* **33**, 289-300 (1985).
- Efron, L., Yeomans, D. K. & Schanzle, A. F. *J. astronaut. Sci.* **33**, 301-323 (1985).
- So, P., Ott, E. & Dayawansa, W. P. *Phys. Lett. A*, (in the press).
- Chen, G. & Dong, X. *From Chaos to Order—Perspectives and Methodologies in Controlling Chaotic Dynamical Systems* Tech. Rep. 92-07 (Univ. Houston, Texas, 1992).
- Vincent, T. L. & Yu, J. *J. Dynam. Control* **1**, 35-52 (1991).
- Grebogi, C., Ott, E. & Yorke, J. A. *Phys. Rev. A* **37**, 1711 (1988).
- Auerbach, D., Cvitanovic, P., Eckmann, J.-P., Gunaratne, G. & Procaccia, I. *Phys. Rev. Lett.* **58**, 2387 (1987).
- Hénon, M. *Commun. Math. Phys.* **50**, 69-77 (1976).
- Ogata, K. *Modern Control Engineering*, 2nd Ed., 347-889 (Prentice-Hall, Englewood Cliffs, 1990).
- Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. *Phys. Rev. Lett.* **45**, 712-716 (1980).
- Eckmann, J.-P. & Ruelle, D. *Rev. mod. Phys.* **57**, 617-656 (1985).
- Whitney, H. *Ann. Math.* **37**, 645-671 (1936).
- Takens, F. in *Lecture Notes in Mathematics* 898 (eds Rand, D. A. & Young, L. S.) 366-381 (Springer, Berlin, 1981).
- Mehta, N. J. & Henderson, R. M. *Phys. Rev. A* **44**, 4861-4865 (1991).
- Schuster, H. G. *Deterministic Chaos*, 2nd ed., 37-38 (VCH, Weinheim, 1989).
- Lorenz, E. *J. Atmos. Sci.* **20**, 130-141 (1963).
- Yorke, J. A., Yorke, E. D. & Mallet-Paret, J. *Physica D* **24**, 279-291 (1987).
- Hayes, S., Grebogi, C. & Ott, E. *Phys. Rev. Lett.* **70**, 3031-3034 (1993).
- Chua, L. O., Komura, M. & Matsumoto, T. *IEEE Trans. Circ. Syst.* **11**, 1073-1118 (1986).
- Hübner, A. & Lüscher, E. *Naturwissenschaften* **76**, 67-69 (1989).
- Hübner, A. W. *Helv. phys. Acta* **62**, 343-346 (1989).
- Plapp, B. B. & Hübner, A. W. *Phys. Rev. Lett.* **65**, 2302-2305 (1990).
- Jackson, E. A. & Hübner, A. W. *Physica D* **44**, 407-420 (1990).
- Jackson, E. A. *Phys. Lett. A* **151**, 478-484 (1990).
- Jackson, E. A. *Phys. Rev. A* **44**, 4839-4853 (1991).
- Lüscher, E. & Hübner, A. W. *Helv. phys. Acta* **62**, 544-551 (1989).
- Georgii, R., Eberl, W. & Lüscher, E. *Helv. phys. Acta* **62**, 290-293 (1989).
- Alekseev, V. V. & Loskutov, A. Y. *Sov. Phys. Dokl.* **32**, 1346-1348 (1987).
- Pettini, M. in *Dynamics and Stochastic Processes* (ed. Lima, R., Streit, L. & Vilela Mendes, R.) 242-250 (Springer, Berlin, 1990).
- Lima, R. & Pettini, M. *Phys. Rev. A* **41**, 726-733 (1990).
- Braiman, Y. & Goldhirsch, I. *Phys. Rev. Lett.* **66**, 2545-2548 (1991).
- Azevedo, A. & Rezende, S. M. *Phys. Rev. Lett.* **66**, 1342-1345 (1991).
- Herzel, H. A. *Angew. Math. Mech.* **68**, 11-12 (1988).
- Fahy, S. & Hamann, D. R. *Phys. Rev. Lett.* **69**, 761-764 (1992).
- Ditto, W. L. et al. *Phys. Rev. Lett.* **63**, 923-966 (1989).
- Lai, Y.-C., Ding, M.-Z. & Grebogi, C. *Phys. Rev. E* **47**, 86-92 (1992).
- Abed, E. H., Wang, H. O. & Chen, R. *Physica D* (in the press).

ACKNOWLEDGEMENTS. This work was supported by the United States Department of Energy (Office of Scientific Computing, Office of Energy Research). We thank D. Auerbach, G. R. Chen, A. Hübner and K. Scarbrough for helpful comments. W. Ditto, S. Raue and M. Spano are gratefully acknowledged for providing data for this article.

Rational design of potent sialidase-based inhibitors of influenza virus replication

Mark von Itzstein*, Wen-Yang Wu*, Gaik B. Kok*, Michael S. Pegg*, Jeffrey C. Dyason*, Betty Jin*, Tho Van Phan*, Mark L. Smythe*, Hume F. White*†, Stuart W. Oliver*, Peter M. Colman‡, Joseph N. Varghese‡, D. Michael Ryan§, Jacqueline M. Woods§, Richard C. Bethell§, Vanessa J. Hotham§, Janet M. Cameron§ & Charles R. Penn§

*Department of Pharmaceutical Chemistry, Victorian College of Pharmacy, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

‡CSIRO Division of Biomolecular Engineering, Parkville, Victoria 3052, Australia

§Glaxo Group Research Ltd, Greenford, Middlesex UB6 0HE, UK

Two potent inhibitors based on the crystal structure of influenza virus sialidase have been designed. These compounds are effective inhibitors not only of the enzyme, but also of the virus in cell culture and in animal models. The results provide an example of the power of rational, computer-assisted drug design, as well as indicating significant progress in the development of a new therapeutic or prophylactic treatment for influenza infection.

INFLUENZA has been with humanity for most of recorded history. Evidence for the potentially lethal nature of the virus dates from as early as 430 BC in the fateful plague of Athens¹. There has been little success in the development of effective anti-influenza virus agents, primarily because of the capacity of the virus to modify its surface antigens, sialidase (EC 3.2.1.18, neuraminidase, acylneuraminyl hydrolase) and haemagglutinin, leaving the immune system unable to cope with essentially new antigens.

The drug amantadine and its analogue rimantidine, are effective in the prophylaxis and treatment of influenza A virus infection. These compounds act by blocking the ion channel function of the virus protein M2 (refs 2,3), which may account for the lack of activity against influenza B viruses, which do not possess this protein. The clinical use of amantadine has been limited by side effects, which may be due to inhibition of endogenous ion channel activity⁴ and to the rapid emergence of resistant viral strains⁵.

The determination of the crystal structure of the two influenza virus surface proteins, haemagglutinin^{6,7} and sialidase⁸⁻¹¹, has provided opportunities for the design of new and specific inhibitors that might have improved potency, and selectivity for the target virus. Sialidase, an enzyme widespread among animals and a number of microorganisms, is a glycohydrolase that cleaves terminal α -ketosidically linked sialic acids from a large array of glycoproteins, glycolipids and oligosaccharides^{12,13}. Many of the sialidase-containing microorganisms are pathogenic in man and it has been proposed that this enzyme plays an important role in the pathogenicity of some of these infections¹⁴⁻¹⁶. The biochemistry of influenza virus sialidase has been investigated¹⁷⁻²¹, and from some of this work a proposal for the mechanism of catalysis for this reaction by sialidases has emerged which involves a sialosyl cation transition-state complex¹⁹.

Since the original reports of an enzyme activity on influenza virus^{22,23} and its further characterization as a sialidase²⁴, it has been postulated that the enzyme helps the elution of newly synthesized virions from infected cells²⁵, and that it may also assist the movement of virus through the mucus within the

respiratory tract^{23,26,27}. Inhibition of this enzyme might therefore result in limiting the establishment and progression of infection by influenza virus, although earlier work by Palese *et al.*²⁸ with the nonselective sialidase inhibitor 2-deoxy-2,3-didehydro-D-N-acetylneuraminic acid (Neu5Ac2en) failed to demonstrate any beneficial effect in animal models of infection.

Influenza sialidase is a tetramer of identical subunits with M_r 60K. The subunit fold is the prototypic β -propeller⁸⁻¹⁰, six four-stranded antiparallel β -sheets arranged as if on the blades of a propeller. An approximate 6-fold symmetry axis passes through the centre of the subunit, relating the six β -sheets and the enzyme active site which lies close to this pseudo-symmetry axis. That site has been identified directly by studying the binding of inhibitors to the enzyme^{8-10,29}. It is a deep cavity on the protein surface and is lined entirely by amino acids that are invariant in sialidases of all strains of influenza virus A and B which have been characterized to date³⁰. Strain-variable amino acids are found next to and encircling the active site⁸⁻¹⁰.

Sialic acid, the product of the enzyme-catalysed reaction, binds to the active site in the boat configuration^{11,29}. The carboxylate, which would be axial and directed to the floor of the active site if the pyranose ring were presenting as a ²C₅ chair conformation, is observed to be equatorial and to interact with three invariant arginine residues on the enzyme. In this conformation it resembles the geometry of the unsaturated sialic acid analogue (Neu5Ac2en), a potent sialidase inhibitor, which is consistent with previous mechanistic studies on influenza virus sialidase^{18,19}. An important aspect of the active site is that strain invariance extends to a number of other amino acids which do not themselves make contact with sialic acid but which rather provide substructure or a scaffold on which amino acids contacting the bound sugar are supported^{11,29}.

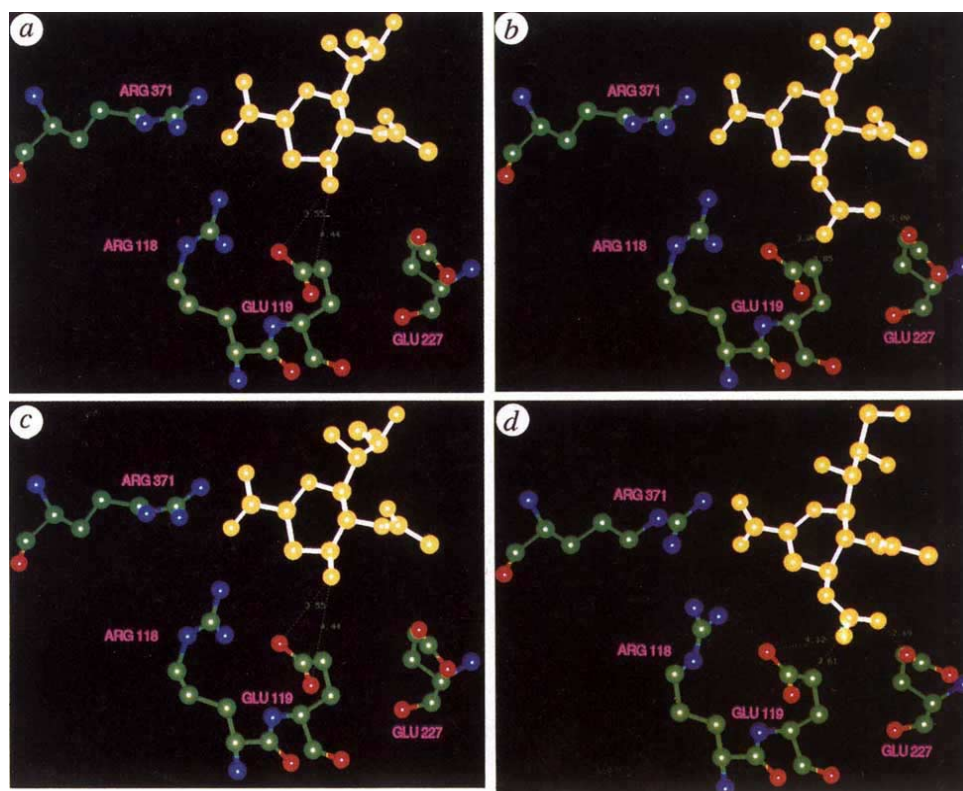
We report here our progress towards the design and evaluation of two sialic acid-based inhibitors of the influenza virus sialidase that are very potent inhibitors of not only the enzyme, but also the virus in cell culture and in animal models.

Design

We have taken several approaches to the use of the crystal structure of influenza virus sialidase in the rational design of inhibitors of this enzyme. A computer-assisted manual examina-

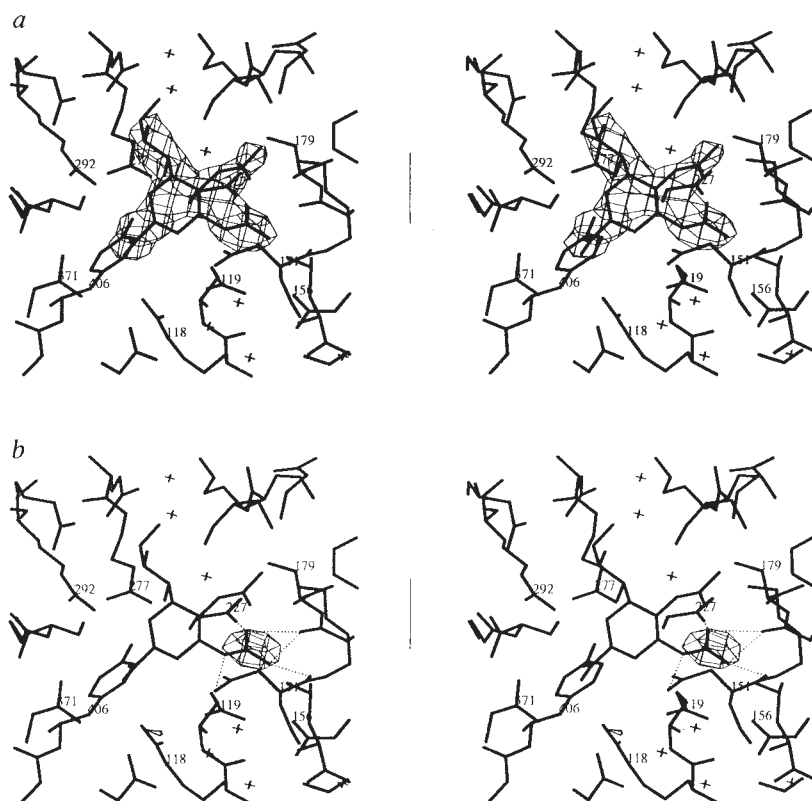
† Present address: CSIRO Division of Wool Technology, Clayton, Victoria, Australia.

FIG. 1 A comparison of the predicted and observed interactions of two inhibitors with influenza virus sialidase. *a*, View of the important amino-acid residues in the influenza sialidase active site predicted to interact with 4-amino-Neu5Ac2en. On the basis of theoretical calculations it is predicted that the 4-amino group of 4-amino-Neu5Ac2en should form a strong salt bridge with the acid group Glu 119, and the increased affinity over the parent template Neu5Ac2en is primarily attributed to the formation of this salt bridge. The salt bridge between Arg 371 and the sialic acid's carboxylic acid group is a common feature for the binding of sialic acid-based sialidase inhibitors and substrates¹⁹. *b*, View of several amino-acid residues within the active site predicted to interact with 4-guanidino-Neu5Ac2en. The 4-guanidinyl group of 4-guanidino-Neu5Ac2en is predicted to interact not only with Glu 119, but also with Glu 227 in a lateral bonding mode with both of the terminal nitrogens of the guanidinyl group³². From our modelling studies the further increase in its affinity for influenza sialidase over 4-amino-Neu5Ac2en can be directly attributed to this additional interaction. *c*, View of the inhibitor 4-amino-Neu5Ac2en bound to the active site of influenza virus A/Tokyo/3/647 sialidase. The structure of the 4-amino-Neu5Ac2en complex is the result of a difference Fourier analysis using a partial (60%) 3 Å dataset only. *d*, View of the inhibitor 4-guanidino-



Neu5Ac2en bound to the active site of influenza virus A/Tokyo/3/67 sialidase. The structure of the 4-guanidino-Neu5Ac2en complex is the result of a crystallographic refinement as described in the text.

FIG. 2 Stereo drawing of the inhibitor 4-guanidino-Neu5Ac2en bound to the active site of influenza virus A/Tokyo/3/67 sialidase. *a*, The overlaid Fourier map is the difference between the 4-guanidino-Neu5Ac2en-sialidase complex and uncomplexed sialidase. The map resolution is 2.8 Å. The contour level of the map is 3σ. *b*, The overlaid Fourier map is of the difference between 4-guanidino-Neu5Ac2en and Neu5Ac2en complexed to the enzyme. The contour level of the map is 4σ, and shows clearly the 4-substituted guanidino group. No significant differences, other than the chemical difference, between the two inhibitor complexes are evident. Amino-acid sequence numbers are indicated and the dotted lines represent hydrogen bonds. The conformation of the carboxylate on the sugar is rotated with respect to the refined structure of the sialic acid complex²⁹, but is likely to be due to the medium resolution (2.8 Å) of the data.



tion of the active site structure, with sialic acid and sialic acid analogues bound, has provided much information^{19,29}. The use of software such as GRID³¹ (a program that determines probable interaction sites between probes with various functional group characteristics and the enzyme surface) to analyse the active site has reinforced our understanding of the interactions of sialic acid and other sialic acid analogues with the enzyme active site. Predictions of energetically favourable substitutions to the already known unsaturated sialic acid analogue Neu5Ac2en were made from these analyses. The most apparent of these was the replacement of the hydroxyl group at the four position on Neu5Ac2en by an amino group. GRID calculations with a protonated primary amine as the probe identified a hotspot around this accessible position at an energy contour level of $-16 \text{ kcal mol}^{-1}$. This result and further inspection of the region suggested that substitution of the 4-hydroxyl group by an amino group should produce a significant increase in the overall binding interaction due to a salt bridge formation with the side chain carboxylic acid group of Glu 119 (Fig. 1a). The replacement of the same hydroxyl group by the significantly more basic guanidinyll group was anticipated to produce an even tighter affinity of the substituted Neu5Ac2en for the active site as a result of lateral binding through the terminal nitrogens of the guanidino group with both Glu 119 and Glu 227 (Fig. 1b). This type of binding mode is well documented and is observed in

a number of protein–ligand complexes³².

The two target molecules were prepared through selective reduction of 4-azido-Neu5Ac2en (ref. 33) to give 4-amino-Neu5Ac2en, and subsequent treatment of this with aminoiminomethanesulphonic acid gave a reasonable yield of 4-guanidino-Neu5Ac2en (ref. 34). Crystallographic studies of 4-amino- and 4-guanidino-Neu5Ac2en complexed to influenza virus A/Tokyo/3/67 sialidase have been done. The first of these (Fig. 1c) is very similar to the parent molecule Neu5Ac2en and the second is described here in detail (Figs 1d and 2). Crystals were soaked for 2 days in 40 mM 4-guanidino-Neu5Ac2en and X-ray diffraction data were collected on a Xenotronics Area Detector (two crystals) and a Mar Image Plate Detector (one crystal) using a rotating anode X-ray source. The 68,500 observations of 21,996 unique reflections (72% of the data to 2.8 Å resolution) merged with an *R*-factor on intensities of 0.12. An atomic model of the inhibitor was built into the difference Fourier map $F_{\text{inhibitor complex}} - F_{\text{native}}$. The model was refined to 2.8 Å resolution in a cycle of simulated annealing³⁵ involving 50 cycles of energy refinement, 1,000 1-fs steps of room temperature dynamics and 50 more cycles of energy refinement. During this refinement all electronic charges on amino-acid side chains were set to zero and only the carboxylate of the inhibitor was formally charged. Harmonic restraints were used for the atom refinement except for atoms within a 12 Å radius of the inhibitor. The *R*-factor for the model is 0.214 and the r.m.s. deviations from the ideal bond lengths and angles are 0.018 Å and 3.7°, respectively.

The model largely agrees with the predicted binding mode for the inhibitors shown in Fig. 1a and b. As anticipated, 4-amino-Neu5Ac2en forms a salt bridge between Glu 119 and the C-4 amino group of 4-amino-Neu5Ac2en (Fig. 1c). For 4-guanidino-Neu5Ac2en it was predicted that a lateral binding mode³² between the terminal guanidinyll nitrogens and the carboxylates of Glu 227 and Glu 119 (Fig. 1b) should be energetically favoured. The observed binding mode for this inhibitor (Fig. 1d) indicates that lateral bonding does occur between Glu 227 and one of the terminal guanidinyll nitrogens, as expected, and that the carboxylate of Glu 119 is within 3.5 Å of the other terminal nitrogen of the guanidinium moiety of the inhibitor, and is presumably still within its electrostatic influence although not as close as predicted.

Inhibition of influenza virus sialidase

Direct measurements of enzyme inhibition have provided further confirmation that the 4-amino- and 4-guanidino-substituted Neu5Ac2en are high-affinity inhibitors for influenza virus sialidase. Using a modification¹⁷ of the fluorometric assay of Potier³⁶, both 4-amino- and 4-guanidino-Neu5Ac2en proved to be potent competitive influenza virus (N2, A/Tokyo/3/67) sialidase inhibitors with inhibition constants (K_i) of 5×10^{-8} and 2×10^{-10} M, respectively.

Furthermore, 4-guanidino-Neu5Ac2en seems to be a slow-binding inhibitor whose apparent K_i changes from 10^{-9} to 10^{-10} M with a half-life of 20 min under our routine assay conditions. This effect is possibly due to a conformational change in either the protein, to better accommodate 4-guanidino-Neu5Ac2en, and/or in the ligand itself upon binding, and is under further investigation. No change in the structure of the protein is evident in the electron density map described above. 4-Guanidino-Neu5Ac2en was evaluated further against sialidases of other influenza virus strains representing both N1 and N2 sialidase subtypes and influenza B virus (Table 1). K_i values, determined by a previously described method³⁷ for evaluation of tight-binding inhibitors, were below 1 nM in each case. The two compounds described here are significantly more potent inhibitors of influenza virus sialidase than any reported previously, including Neu5Ac2en and the *N*-trifluoroacetylated analogue with K_i s of 1×10^{-6} M (under our assay conditions) and 8×10^{-7} M (ref. 38), respectively.

TABLE 1 Inhibition of influenza virus sialidases by 4-guanidino-Neu5Ac2en

Virus	K_i (M)
A/Brazil/11/78 (N1)	$0.75 \times 10^{-10} \pm 0.26 \times 10^{-10}$
Reassortant X31 (N2)	$1.50 \times 10^{-10} \pm 0.55 \times 10^{-10}$
B/Hong Kong/3/91	$7.0 \times 10^{-10} \pm 2.2 \times 10^{-10}$

All enzyme reactions were done in 32.5 mM 2-(*N*-morpholino)-ethane sulphonic acid, 40 mM calcium chloride using intact virions as the source of enzyme and 2'-(4-methylumbelliferyl)-*N*-acetylneuraminic acid as the substrate in a final volume of 350 µl. After preliminary experiments to determine the K_m of the substrate with each enzyme, and the time-dependent nature of the inhibition by 4-guanidino-Neu5Ac2en, K_i values were determined by the method of Henderson³⁷. 4-Guanidino-Neu5Ac2en, or blank where appropriate, and enzyme were preincubated for 10 min in buffer. Reactions were initiated by the addition of substrate, five 50-µl aliquots were removed over a 15-min time course and each was immediately quenched by addition into 150-µl aliquots of 0.014 M sodium hydroxide in 83% aqueous ethanol in black microFLUOR 96-well plates. The concentration of product, 4-methylumbelliferone, was determined fluorimetrically using a Perkin-Elmer-LS-5B luminescence spectrometer equipped with a 96-well plate reader using an excitation wavelength of 365 nm and an emission wavelength of 450 nm. Plots of fluorescence against time, which were essentially linear over the 15-min assay period, were used to obtain values for the initial velocity in the presence (v_i) or absence (v_0) of inhibitor. Because error is present in both axes in a Henderson plot, making it inappropriate for standard regression analysis⁴¹, K_i values were obtained from the data by fitting to a modified version of the Henderson equation for competitive inhibition:

$$Qr^2 + (E_t - Q - I_t)r - E_t = 0$$

where, using the notation of Henderson³⁷, $Q = K_t \left(\frac{A_t + K_a}{K_a} \right)$ and, $r = \frac{v_0}{v_i}$

This equation was solved for the positive root with the constraint that $Q = K_t((A_t + K_a)/K_a)$ using PROCNLIN from SAS (SAS Institute Inc., Cary, North Carolina, USA) which performed nonlinear regression using least-square techniques. The iterative method used was the multivariate secant method, similar to the Gauss–Newton method except that the derivatives in the Taylor series are estimated from the histogram of iterations rather than supplied analytically. The convergence criterion used was a change in loss function of less than 10^{-8} .

The selectivity of 4-amino- and 4-guanidino-Neu5Ac2en for influenza virus sialidase is being investigated. Our preliminary *in vitro* results show that these compounds are less active against mammalian sialidases such as human placental sialidase and partially purified sheep liver sialidase, bacterial sialidases and para-influenza sialidase (Table 2). The affinities of these compounds for human placental sialidase seem to be about one million times lower than is observed with influenza virus N2 sialidase. In addition, we have investigated the inhibition of sialidase activity from sheep liver, *Arthrobacter ureafaciens*, *Clostridium perfringens*, *Vibrio cholerae* and parainfluenza virus by a range of eight Neu5Ac2en analogues modified at the C-4 position, including 4-amino- and 4-guanidino-Neu5Ac2en. All of these C-4 modified analogues are weaker inhibitors of these sialidases than Neu5Ac2en³⁹, in contrast to the results obtained for influenza virus. Neu5Ac2en does not discriminate between viral, bacterial and mammalian sialidases. These observations suggest that the active site architecture and chemistry of influenza virus sialidase, in particular the nature of the C-4-binding pocket, is different to these other sialidases examined so

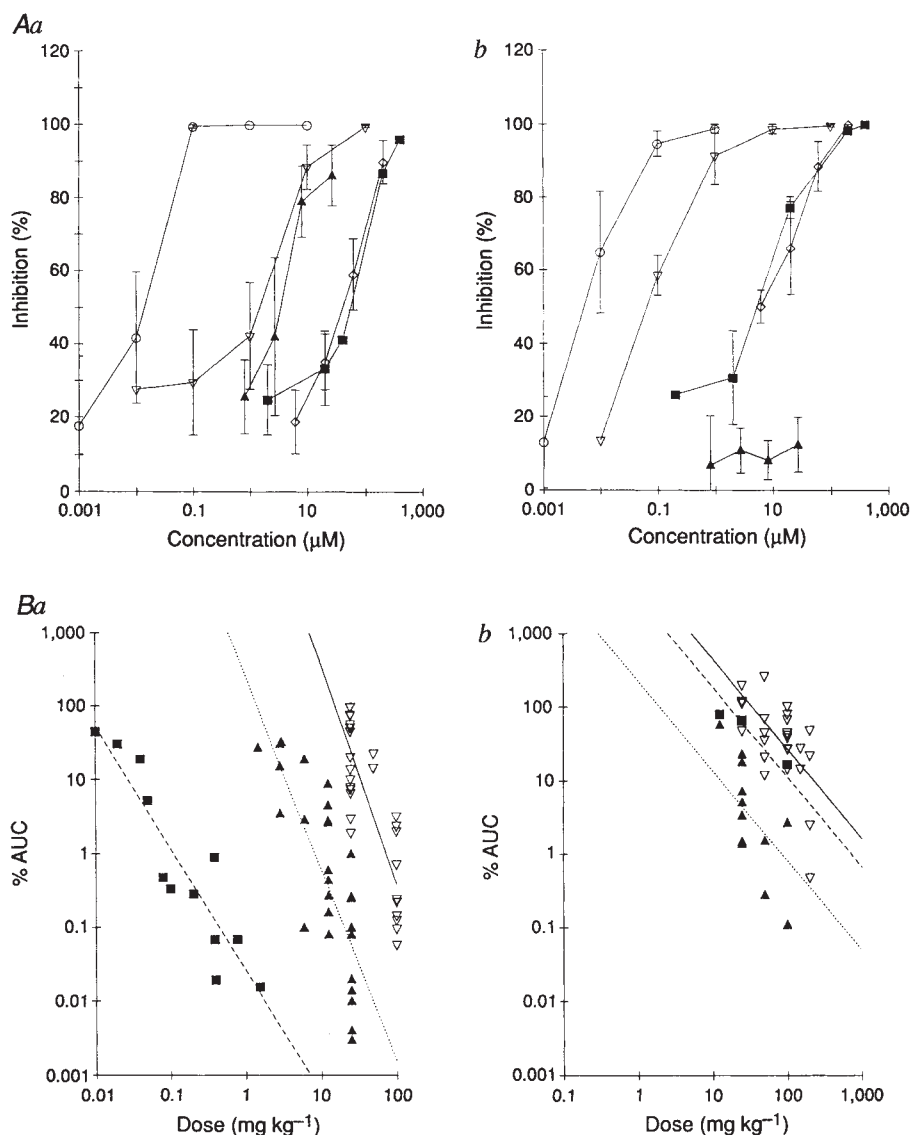
far³⁹ and that inhibitor design based on the target structure can result in highly selective compounds.

Inhibition of influenza virus replication

The results described above clearly demonstrate that the design of tight-binding inhibitors of the influenza virus sialidase, based on the crystal structure of this enzyme, is possible. Equally important is the confirmation of the potential of computer software such as that developed by Goodford³¹ in prediction of 'logical' substitutions on the sialic acid-based template. But demonstration of clinically useful levels of inhibition of virus replication is also important if this approach is to be of practical benefit.

Both 4-amino- and 4-guanidino compounds are effective inhibitors of influenza A (Singapore/1/57) and B (Victoria/102/85) viruses in tissue culture, with IC₅₀ values (the concentration required to reduce plaque formation in MDCK cells by 50%) of 1.5 μ M and 0.065 μ M (4-amino) and 0.014 μ M and 0.005 μ M (4-guanidino), respectively (Fig. 3A). The 4-guanidino-Neu5Ac2en in particular shows dramatic inhibition of influenza

FIG. 3A Inhibition of influenza viruses A/Singapore/1/57 (a) and B/Victoria/102/85 (b) plaque formation by 4-guanidino-Neu5Ac2en (○), 4-amino-Neu5Ac2en (▽), Neu5Ac2en (■), amantadine (▲) and ribavirin (◇). Plaque assays were done essentially as described in ref. 43. MDCK cells were inoculated with influenza A/Singapore/1/57 (a) or influenza B/Victoria/102/85 (b) diluted in Eagle's minimal essential medium (pH 7.2–7.4) containing 2 μ g ml⁻¹ trypsin to give ~60 plaques per well. Cells were left for 1 h at room temperature for virus to adsorb, and then overlaid with cell growth medium (DCCM-1, Biological Industries) containing 0.5% agarose, 2 μ g ml⁻¹ trypsin, 0.001% DEAE dextran and test compound. After 72 h at 37 °C plaques were visualized by fixing with 5% glutaraldehyde and staining with carbol fuchsin. The percentage inhibition of plaque formation relative to controls (plaques formed in absence of any inhibitor) were calculated for each inhibitor concentration. The means of per cent inhibition values from three experiments were used to determine the IC₅₀ for each inhibitor. B, Inhibition of influenza virus A/Singapore/1/57 replication in mice by 4-guanidino-Neu5Ac2en (■, ---), amantadine (▲,), and ribavirin (▽, ---). Female CD-1 mice were infected intranasally (a) with 1.5 $\times$ 10² TCID₅₀ of influenza A/Singapore/1/57. Compounds were administered 18 and 3 h before infection, 3 h after and then twice daily on days 1–3 after infection. For intranasal administration mice were anaesthetized by inhalation of ether and 50 μ l of compound in PBS placed into the nasal passages. Intraperitoneal administration (b) was by injection of 0.5 ml of compound in PBS. Control animals were dosed with PBS alone. Four animals from each group were culled on days 1–3 postinfection and virus in lung homogenates titrated by tissue culture infectivity. The mean virus titres (TCID₅₀) for days 1–3 were plotted against time, and the area under the curve calculated. These values, as a percentage of the corresponding control (%AUC), were plotted against dose. Lines were fitted to these data (using SAS programs, SAS Institute Inc.) on the assumption that the dose responses, on log scales, are linear, and the ED₅₀ for each compound is taken as the dose corresponding to a %AUC of 10 using these fitted lines. A direct comparison of the efficacies of different inhibitors is only possible where the dose



responses for each compound within a given treatment are the same. A model was therefore used that gave a fit of parallel lines to each graph but the slopes shown in (a) were significantly different ($P < 0.019$) precluding the use of a parallel lines model. The ED₅₀ values in (a) were therefore compared using an approximate *t*-test.

virus in cell culture, with IC_{50} values well below those of amantadine, ribavirin and Neu5Ac2en (Fig. 3A), indicating that the potent inhibition seen against isolated enzyme does correlate with inhibition of virus replication.

Previous attempts to demonstrate clinical potential of inhibitors of the influenza virus sialidase have been unsuccessful²⁸, despite the compounds (Neu5Ac2en and its *N*-trifluoroacetyl analogue) having comparable activity to amantadine and ribavirin against influenza virus in cell culture assays. In those studies²⁸ there was no effect on viral replication after intranasal and subcutaneous treatment of mice, suggesting that the Neu5Ac2en-based inhibitors might not be suitable candidate drugs for treatment of influenza infection.

Nöhle *et al.*⁴⁰ subsequently reported that the excretion of Neu5Ac2en from the mouse is very rapid and suggested that this could be the reason for the apparent lack of activity of this

compound in influenza-infected mice. In evaluating our sialidase inhibitors for anti-viral activity in animal models (both mouse and ferret) we have concentrated on an intranasal route of administration, because it is believed that the virus replicates in surface respiratory epithelial cells, and that the sialidase has primarily an extracellular role. Using this approach we have demonstrated inhibition of virus replication in both mice and ferrets, with both 4-amino- and 4-guanidino-Neu5Ac2en and, to a lesser extent, with Neu5Ac2en itself. The most comprehensive data have been obtained with 4-guanidino-Neu5Ac2en in mice infected with influenza A/Singapore/1/57, in comparison with amantadine and ribavirin (Fig. 3B); efficacy of the inhibitors in this model is expressed as an EDAUC10, that is the dose required to reduce titres of virus in lung homogenates to 10% of that in control (infected) animals. When administered intranasally to mice infected with influenza A virus, 4-guanidino-

FIG. 4 Treatment of influenza-infected ferrets with 4-guanidino-Neu5Ac2en. *a*, Influenza virus titres measured in nasal washings from ferrets treated with 4-guanidino-Neu5Ac2en (●) or distilled water (■). *b*, Body temperature recordings for each of the 5 ferrets (1 line per animal) treated with 4-guanidino-Neu5Ac2en (●) or distilled water (■). Ferrets (0.7–1.3 kg) were infected intranasally with 10^4 TCID₅₀ influenza A/Mississippi/1/85. Groups of 5 animals were dosed intranasally with either 4-guanidino-Neu5Ac2en or distilled water (0.25 ml kg^{-1}) twice daily (8 h apart) from the day before infection to day 5 after infection. On the day of infection the first treatment dose was given 2 h before infection. To help intranasal dosing, the ferrets were sedated with gaseous isoflurane (Abbot Lab Ltd, UK). Infection was monitored by taking nasal wash samples on days 1–9 after infection, and virus titres determined as tissue culture infectivity (in MDCK cells) by limiting dilution. Values are shown as the mean from 5 animals, $\pm$ s.e.m. Temperature responses were monitored by telemetry. Battery operated transmitters (model TA10TA-F50 special temperature implants, Data Sciences Inc., USA) were implanted into the peritoneal cavity 7 days before infection. Telemetric data were collected by dedicated receivers (RLA 2000, Data Sciences Inc., USA) for analysis and conversion to a graphical format using the Dataquest IV system (Data Sciences Inc., USA). Influenza-induced pyrexia was detected over days 3–5 with peak temperature (39.5°C) occurring on days 3.5–4.5; baseline temperatures were in the region of 37.5 – 38°C .

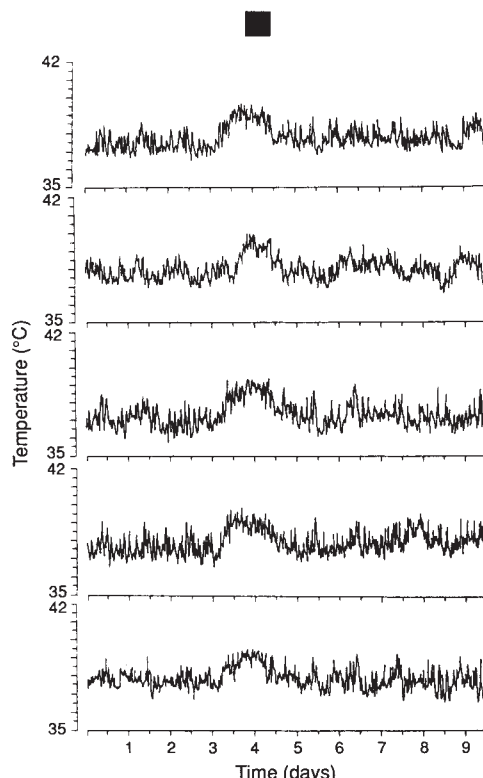
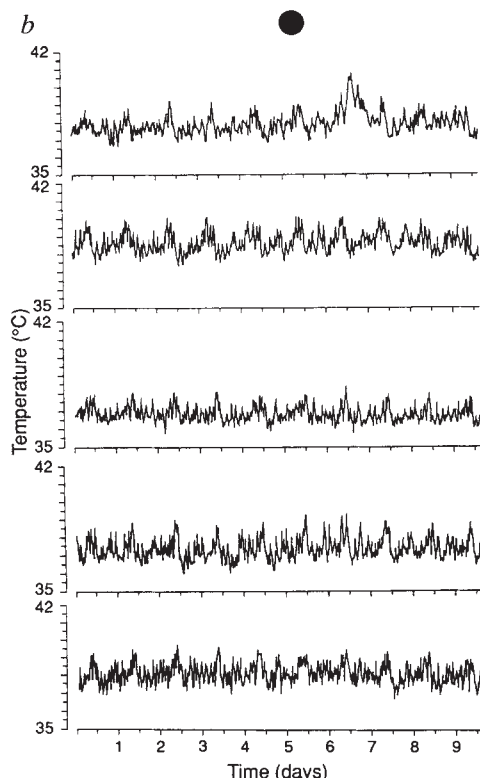
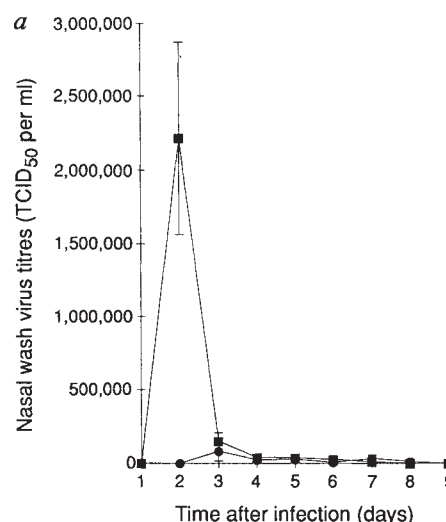


TABLE 2 Inhibition of other sialidases

Compound	Sialidase			
	IC ₅₀ (M)* Human	Paraflu	K _i (M)† <i>V. chol.</i>	Sheep
Neu5Ac2en	1.2 × 10 ⁻⁵	1.0 × 10 ⁻⁵	3.0 × 10 ⁻⁶	8.0 × 10 ⁻⁶
4-Amino- Neu5Ac2en	9.3 × 10 ⁻³	3.0 × 10 ⁻⁴	3.0 × 10 ⁻⁴	1.0 × 10 ⁻⁴
4-Guanidino- Neu5Ac2en	1.0 × 10 ⁻³	8.0 × 10 ⁻⁴	6.0 × 10 ⁻⁵	3.0 × 10 ⁻⁴

Human lysosomal sialidase (human) was purified as a complex with β -galactosidase and carboxypeptidase from placenta by the method of Potier *et al.*⁴². Parainfluenza 2 virus (paraflu), *Vibrio cholerae* (*V. chol.*) and sheep liver sialidases (sheep) were isolated or purchased as reported previously³⁹.

*All human lysosomal sialidase assays were done at 37 °C using 4-methylumbelliferyl-N-acetylneuraminic acid as substrate in 0.1 M sodium acetate buffer, pH 4.3. The K_m for this substrate was determined to be 85 ± 5 μ M. Inhibitor concentrations (M) causing 50% reduction of human lysosomal sialidase activity (IC₅₀) were determined by measuring initial velocities at varying inhibitor concentration at a fixed concentration of substrate (85 μ M), and were obtained from plots of fractional inhibition against log₁₀[inhibitor]. IC₅₀ values were determined rather than K_i values because of the small amounts of protein available. For a competitive inhibitor, the IC₅₀ value determined at [Substrate] = K_m will have the value 2 × K_i .

†Assays and K_i determinations for sialidases from paraflu, *V. chol.* and sheep were as described previously³⁹.

Neu5Ac2en had an EDAUC10 of 0.027 mg kg⁻¹, which was much lower than the corresponding values for amantadine (3.51 mg kg⁻¹) and ribavirin (33.9 mg kg⁻¹) (Fig. 3*Ba*). By contrast, when 4-guanidino-Neu5Ac2en was administered to infected mice intraperitoneally, rather than intranasally, it was far less effective (Fig. 3*Bb*). The EDAUC10 values for amantadine and ribavirin were three to sevenfold higher (12.8 mg kg⁻¹ and 223 mg kg⁻¹, respectively) whereas 4-guanidino-Neu5Ac2en was ~4,000-fold less active (EDAUC10 of 108 mg kg⁻¹) when given intraperitoneally compared with intranasally.

Results obtained with 4-guanidino-Neu5Ac2en administered both before and during infection of ferrets at a dose of 50 μ g kg⁻¹ twice daily (b.d.) are shown in Fig. 4. In this experimental

system, virus shedding is observed in nasal washes from 3 to 5 days postinfection, and an elevated body temperature recorded on day 2–3 postinfection. At 50 μ g kg⁻¹ (b.d.) 4-guanidino-Neu5Ac2en effectively eliminated virus shedding and reduced pyrexia fully in all 5 animals. 4-Amino-Neu5Ac2en showed a similar effect in this system at ~6 mg kg⁻¹ (b.d.). 4-Guanidino-Neu5Ac2en was ~1,000 times more effective than amantadine when given intranasally to either species.

The question of resistant mutants arising under pressure from these sialic acid analogues, not unlike those that have emerged for amantadine³, does arise. Both 4-amino- and 4-guanidino-Neu5Ac2en bind entirely within the conserved catalytic site of the enzyme, and make no contacts with strain variable residues outside that site. But 4-guanidino-Neu5Ac2en, by virtue of the extra bulk now at the four position, establishes contact with some of the structural amino acids of the active site, that is, those referred to above which themselves make no direct contact with sialic acid or unsubstituted Neu5Ac2en. For example, Glu 227 is hydrogen bonded to the guanidino moiety of the inhibitor (Fig. 1*d*) but makes no contact with either sialic acid or Neu5Ac2en²⁹. Similarly, Glu 119, which may assist in balancing the charge on the 4-guanidino group, makes no contact with bound sialic acid. There exists a formal possibility that viral mutants might emerge that could take advantage of the differential binding properties of the substrate and the inhibitor to allow selectively the binding of the former but not the latter. In our preliminary experiments in animals, under conditions which rapidly give rise to amantadine resistance in patients being treated with that drug, we have failed to isolate resistant virus under the pressure of 4-guanidino-Neu5Ac2en in the laboratory. Further work is in progress to see if such variants can be selected.

In conclusion, the two target compounds (4-amino- and 4-guanidino-Neu5Ac2en) are the most potent influenza virus sialidase inhibitors reported. Their inhibitory activity on the enzyme extends to both *in vitro* and *in vivo* activity against influenza virus replication. Importantly, these compounds have similar activity against both influenza A and B. The significant *in vivo* activity and the apparent specificity of these two compounds against influenza suggest that they may be useful leads for the development of novel anti-influenza drugs and further studies are in progress. Moreover, we also consider that these results provide an example of the usefulness of rational, computer-assisted drug design based on the crystal structure of a target protein. □

Received 1 September 1992; accepted 14 April 1993.

- Langmuir, A. D., Worthen, T. D., Solomon, J., Ray, C. G. & Petersen, E. *New Engl. J. Med.* **313**, 1027–1030 (1985).
- Hay, A. J., Wolstenholme, A. J., Skehel, J. J. & Smith, M. H. *EMBO J.* **4**, 3021–3024 (1985).
- Pinto, L. H., Holsinger, L. J. & Lamb, R. A. *Cell* **69**, 517–528 (1992).
- Stoof, J. C., Boonij, J., Drukarch, B. & Wolters, E. C. *Eur. J. Pharmacol.* **213**, 439–443 (1992).
- Hayden, F. G. *et al.* *New Engl. J. Med.* **321**, 1696–1702 (1989).
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
- Wilson, D. C., Wilson, I. A. & Skehel, J. J. *Nature* **289**, 373–378 (1981).
- Varghese, J. N., Laver, W. G. & Colman, P. M. *Nature* **303**, 35–40 (1983).
- Colman, P. M., Varghese, J. N. & Laver, W. G. *Nature* **303**, 41–44 (1983).
- Varghese, J. N. & Colman, P. M. *J. molec. Biol.* **221**, 473–486 (1991).
- Burmeister, W. P., Ruigrok, R. W. H. & Cusack, S. *EMBO J.* **11**, 49–56 (1991).
- Klenk, E., Faillard, H. & Lempfrid, H. *Hoppe-Seyler's Z. physiol. Chem.* **301**, 235–246 (1955).
- Gottschalk, A. *Biochim. biophys. Acta* **23**, 645–646 (1957).
- Drzeniek, R. *Curr. Top. microbiol. Immun.* **59**, 35–74 (1972).
- Schauer, R. *Trends biochem. Sci.* **10**, 357–360 (1985).
- Schauer, R. in *Structural Carbohydrates in the Liver* (eds Popper, H., Reutter, W., Kottgen, E. & Gudat, F.) 83–97 (MTP, Boston, 1983).
- Chong, A. K. J., Pegg, M. S. & von Itzstein, M. *Biochim. biophys. Acta* **1077**, 65–71 (1991).
- Chong, A. K. J., Pegg, M. S. & von Itzstein, M. *Biochem. Int.* **24**, 165–171 (1991).
- Chong, A. K. J., Pegg, M. S., Taylor, N. R. & von Itzstein, M. *Eur. J. Biochem.* **207**, 335–343 (1992).
- Dimmock, N. J. *J. gen. Virol.* **13**, 481–483 (1971).
- Carroll, S. M. & Paulson, J. C. *Arch. Virol.* **71**, 273–277 (1982).
- Hirst, G. K. *J. exp. Med.* **76**, 195–209 (1942).
- Burnet, F. M. *Aust. J. exp. Biol. med. Sci.* **26**, 381–387 (1948).
- Gottschalk, A. *Biochim. biophys. Acta* **23**, 645–646 (1957).
- Palese, P., Jobita, K., Ueda, M. & Compans, R. W. *Virology* **61**, 397–410 (1974).

- Colman, P. M. & Ward, C. W. *Curr. Top. microbiol. Immun.* **114**, 117–255 (1985).
- Klenk, H.-D. & Rott, R. *Adv. Virus Res.* **34**, 247–280 (1988).
- Palese, P. & Schulman, J. L. in *Chemoprophylaxis and Virus Infections of the Upper Respiratory Tract Vol. 1* (ed. Oxford, J. S.) 189–205 (CRC, Cleveland, 1977).
- Varghese, J. N., McKimm-Breschkin, J., Caldwell, J. B., Kortt, A. A. & Colman, P. M. *Proteins* **14**, 327–332 (1992).
- Colman, P. M. in *The Influenza Viruses* (ed. Krug, R. M.) 175–218 (Plenum, New York, 1989).
- Goodford, P. J. *J. med. Chem.* **28**, 849–857 (1985).
- Tintelnot, M. & Andrews, P. J. *Comp.-Aided. molec. Design* **3**, 67–84 (1989).
- von Itzstein, M., Jin, B., Wu, W.-Y. & Chandler, M. *Carbohydr. Res.* (in the press).
- von Itzstein, M., Wu, W.-Y., Phan, T. V., Danyelec, B. & Jin, B. *Chem. Abs.* **117**, 49151y (1991).
- Brunger, A. T. *J. molec. Biol.* **203**, 803–816 (1988).
- Potier, M., Mameli, L., Belisle, M., Dallaire, L. & Melancon, S. B. *Analyt. Biochem.* **94**, 287–296 (1979).
- Henderson, P. J. F. *Biochem. J.* **127**, 321–333 (1972).
- Meindl, P., Bodo, G., Palese, P., Schulman, J. L. & Tuppy, H. *Virology* **58**, 457–463 (1974).
- Holzer, C. T. *et al.* *Glycoconjugate J.* **10**, 40–45 (1993).
- Nöhle, U., Beau, J.-M. & Schauer, R. *Eur. J. Biochem.* **126**, 543–548 (1982).
- Leatherbarrow, R. J. *Trends biochem. Sci.* **15**, 455–458 (1990).
- Potier, M., Michaud, L., Tranchemontagne, J. & Thauvette, L. *Biochem. J.* **267**, 197–202 (1990).
- Hayden, F. G., Cote, K. M. & Douglas, R. G. Jr *Antimicrob. Ag. Chemother.* **17**, 865–870 (1980).

ACKNOWLEDGEMENTS. We thank B. Danyelec, B. Mackey, M. Radovanovic, F. Rose, C. Holzer, M. Trezise, W. Stewart, A. van Donkelaar, J. Tilling, B. A. Pearson, S. A. Hiscox, M. Dempsey and J. Ticehurst for technical assistance and G. E. Amphlett and E. A. Harrison for assistance with analysis of data. We thank the Australian Government, through an Industry Research and Development Board generic grant and the National Health & Medical Research Council, and Biota Holdings Pty Ltd and Glaxo, Australia for financial support.

Radio observations of the classical nova Cygni 92 eighty days after outburst

P.E. Pavelln*, R.J. Davis*, L.V. Morrison†, M.F. Bode‡ & R.J. Ivson§

*University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield SK11 9DL, UK

†Royal Greenwich Observatory, Madingley Road, Cambridge CB3 0EZ, UK

‡Department of Chemical and Physical Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

§Department of Astronomy, University of Toronto, 60 St George Street, Toronto, Ontario, M5S 1A7, Canada

CATAclysmic variables are interacting binary stellar systems containing a main-sequence secondary component and a collapsed primary component, usually a white dwarf. A subset of these systems give rise to classical novae, in which an explosive outburst caused by 'runaway' thermonuclear processes on the surface of the white dwarf produces a large increase in visual magnitude and the ejection of substantial quantities of stellar material¹. Radio emission has been detected from several classical novae¹⁷, and is attributed to the thermal bremsstrahlung mechanism operating within an isothermal, expanding shell of ionized gas. Such observations offer a means of directly probing the shells during the early stages of their evolution — long before they can be resolved by optical telescopes. Here we report radio observations of one of the brightest novae in recent years, Nova Cygni 92, using the MERLIN radiotelescope array. The structure of the shell is first resolved eighty days after the initial outburst, revealing nonspherical expansion and thermal electron temperatures that rapidly exceed the expected value for a radiatively excited hydrogen plasma. We explain the observed temperatures by means of a model in which the gas is heated in shocks produced by the thermonuclear runaway.

The binary separation of the two component stars in a cataclysmic variable is such that the secondary fills its Roche lobe and matter flows through the inner lagrangian point of the binary system to the primary companion. This matter is accreted onto the primary through either a disk or a column, depending on how strongly the accretion flow is constrained by the white dwarf's magnetic field. In the case of classical novae, the accretion process culminates in an explosive outburst which gives a large rise in visual magnitude accompanied by the ejection of 10^{-5} to 10^{-4} solar masses ($M_{\odot}$) of matter. The outburst is thought to be caused by thermonuclear runaway in a degenerate envelope of hydrogen-rich gas which has built up on the surface of the white dwarf after $\sim 10^4$ yr of accretion¹.

Nova Cygni 92 was first reported² at a visual magnitude of 6.8 on 19 February 1992 rising to magnitude 4.4 two days later, making it one of the brightest novae of recent years. The development of the optical light-curve shows Nova Cygni 92 to be a fast nova³, and spectroscopic observations in the optical, infrared and ultraviolet show the nova to be evolving rapidly, with complex, numerous absorption features. The presence of enhanced line emission from neon shows the source to be a probable ONeMg nova⁴. The nova, resolved⁵ using the Naval Research Laboratory Mark III optical interferometer on Mount Wilson, had an angular diameter of ~ 4 mas 10 days after outburst. The first radio detection, by the Very Large Array (VLA), shows the nova to have an optically thick, thermal spectrum⁶.

The optical position of the nova was observed with MERLIN (the Multi-Element Radio Linked Interferometer Network) at four separate epochs (9 May, 11 June, 22 July and 9 August 1992) for a period of 24 hours each. The full MERLIN array was

used with a bandwidth of 30 MHz at a frequency of 5 GHz using cooled High Electron Mobility Transistor receivers on six telescopes. The maximum baseline length was 218 km, giving a resolution of 50 mas. The observations were made using a phase-referencing technique in which 7-min observations of Nova Cygni 92 were interspersed with 1-min observations of the nearby very-long-baseline interferometer (VLBI) point-source calibrator 2031+549. The angular separation of the target and reference sources is $\sim 2^{\circ}$. The absolute flux-density scale was set with respect to an observation of 3C286 which was assumed to be unresolved on the shortest baselines; its 5-GHz flux density was taken to be 7.31 Jy (ref. 7). All data were analysed using the Jodrell Bank SERC STARLINK node. Initial data reduction was done with the local OLAF package and the data were mapped using standard phase-referencing techniques in the National Radio Astronomy Observatory Astronomical Image Processing System (AIPS). Complex telescope gain solutions were derived by applying self-calibration to the phase reference source only, and these solutions were applied to the target source. The data were then mapped and CLEANed⁸ to produce the final images, which have a synthesized clean beam diameter of 50 mas full width at half maximum (see Fig. 1). The noise level in each map was measured to be $50 \mu\text{Jy beam}^{-1}$, close to the design goal of the new instrument.

A radio source was detected⁹ about $2''$ away from both the targeted optical position and that of a suggested ~ 18 -mag precursor object¹⁰ visible on the Palomar Sky Survey (PSS). Subsequently a more accurate optical position of the nova was measured using the CAMC (Carlsberg Automated Meridian Circle) on La Palma; the mean position derived from 22

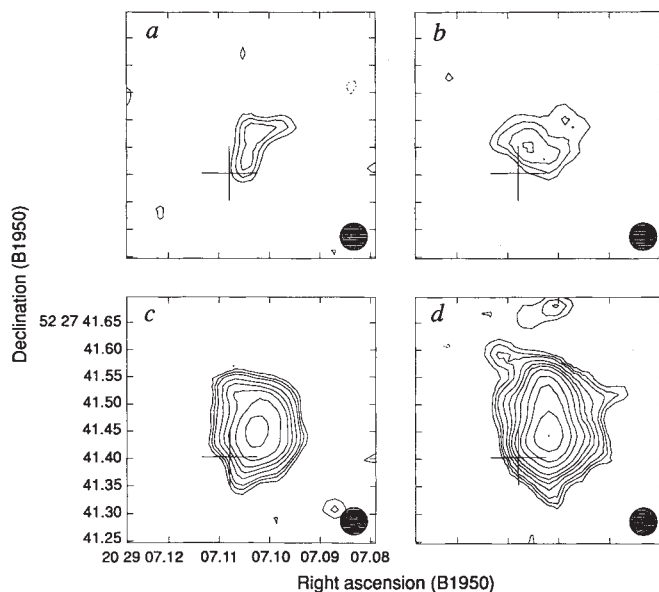


FIG. 1 MERLIN maps of Nova Cygni 1992 at 5 GHz. *a*, Epoch 9 May; *b*, 11 June; *c*, 22 July; *d*, 9 August. The peak brightnesses are 310, 360, 1,140 and 2,140 $\mu\text{Jy beam}^{-1}$ and the integrated flux densities are 1.4, 2.0, 7.4 and 11.5 mJy respectively. The resolution is 50×50 mas. The contours are at equal logarithmic intervals of a factor 1.3 above a bottom contour of $150 \mu\text{Jy beam}^{-1}$. Negative (dashed) contours are at $-150 \mu\text{Jy beam}^{-1}$. The beam size (full width at half maximum) is indicated in the bottom right of each map. The bottom left tick marks are right ascension $\alpha = 20^{\text{h}} 29^{\text{m}} 07.12^{\text{s}}$ and declination $\delta = 52^{\circ} 27' 41.25''$ (B1950), with tick intervals of 0.01 s and $0.05''$ respectively. The cross indicates the optical position measured by CAMC from 22 observations. The size of the cross indicates the optical 1σ error of 50 mas. North is to the top of the images and east is to the left. Small wiggles around the periphery of the images are consistent with noise.

individual measurements was: $\alpha = 20\text{ h } 30\text{ min } 31.662\text{ s}$, $\delta = +52^\circ 37' 50.83'' (\pm 50\text{ mas})$. The mean position of the radio peak from the four epochs was $\alpha = 20\text{ h } 30\text{ min } 31.6570\text{ s}$, $\delta = +52^\circ 37' 50.884'' (\pm 10\text{ mas})$. Both positions are equinox J2000, epoch 1992.5. These positions, indicated in Fig. 1, are consistent within the quoted errors, thus confirming that the observed radio emission is from the nova and furthermore that the nova is not associated with the PSS object. Because the PSS object and the nova are so close, however, a true precursor object as bright as $\sim 20\text{ mag}$ would be undetectable on the PSS plate, so the nova must have risen by at least 16 mag at outburst.

Figure 1a–d shows maps of the radio source at each of the four epochs, together with the CAMC position and its associated 1σ error. The results show that the source was initially extended in the north–south direction; by the second epoch, it had expanded along the minor axis and become less elongated. The total flux density in the source rose from 1.4 mJy at the first epoch to 2.0 mJy at the second. By the third epoch, the total flux density had risen substantially to 7.4 mJy and the ejecta appeared to be more spherically symmetric, although a north–south extension persisted. The flux density continued to rise rapidly to 11.5 mJy at the fourth epoch. There is some extension to the northwest in the first two epochs and to the northeast in the last two. The maps are consistent with the peak of the radio emission being coincident with the optical nova. The radio light-curve (Fig. 2a) shows the total 5 GHz flux density to increase slowly at first (May–June epochs), then more rapidly (June–August). At the latter epochs, the flux density is increasing at a rate faster than the t^2 proportionality (where t is time since outburst) that one normally associates with an optically thick, isothermal expanding shell. This behaviour is accompanied by a similar trend in the peak brightness temperature T_b from the maps: it is initially a roughly constant $\sim 10^4\text{ K}$ (the canonical value for a radiatively excited hydrogen plasma), but rises rapidly to $> 6 \times 10^4\text{ K}$ by the last epoch. As the optical depth $\tau > 1$ (R. M. Hjellming, personal communication), T_b is a true measure of the electron temperature of the gas. This rise in brightness temperature, together with the physical expansion, explains the $> t^2$ behaviour of the radio light curve. JCMT observations¹¹ at millimetre and submillimetre wavelengths, combined with the 5 GHz flux density at a similar epoch, show an optically thick thermal spectrum with a frequency dependence of $\nu^{1.5}$. This is the highest thermal emission brightness temperature measured for a classical nova, and such high temperatures are difficult to produce by radiative heating. We therefore propose that the gas is heated in shocks, directly or indirectly, produced by the thermonuclear runaway. A similar temperature was inferred from a short-lived flare on Nova Vulpeculae 84 No.2 (ref. 12), but this has been attributed to nonthermal emission. The recurrent nova RS Ophiuchi¹³ displayed temperatures of up to 10^8 K but this is thought to have been due to strong shock interaction between the ejecta and a pre-outburst envelope.

The interferometer map contours are in units of mJy per beam which by the Rayleigh–Jeans law transform directly to T_b . Thus by following identically labelled contours, it is possible to track the motion of regions with the same brightness temperature from epoch to epoch. Any adiabatic cooling will, however, result in an underestimate of the expansion rates. Also, for regions $\leq 50\text{ mas}$ in size, resolution effects will bias up the measured sizes and bias down the measured brightness temperatures. It is evident that the structure cannot be fitted by two-dimensional gaussian components. The north–south and east–west angular diameters of regions of different T_b are illustrated by Fig. 2b and c as a function of epoch. The graphs show that all regions expanded monotonically except for the lowest temperature region which initially showed no expansion in the north–south direction. We note that the graphs do not point back towards the origin, indicating that regions of successively hotter gas are becoming visible at correspondingly later times after outburst. This may be explained in terms of a

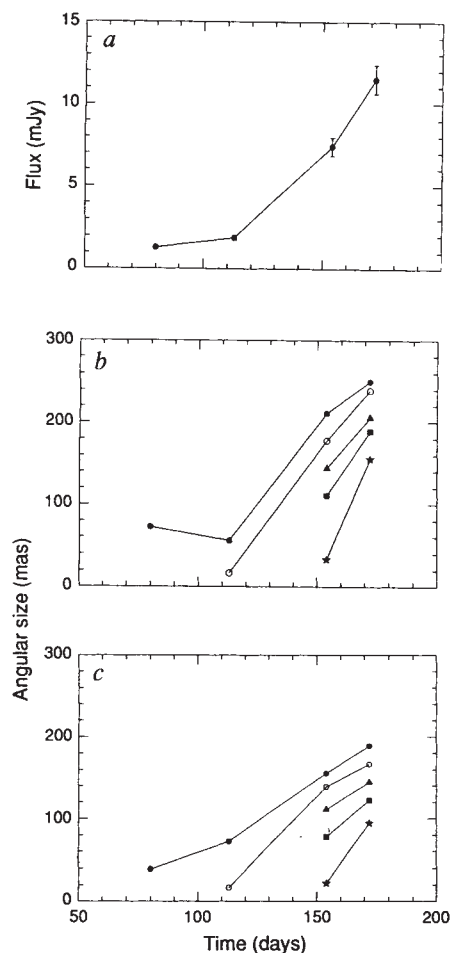


FIG. 2 Radio light curve and spatial expansion of Nova Cygni 92. *a*, Radio light curve at 5 GHz from 80 to 172 days after the optical outburst, with 1σ error bars. *b*, The north–south and (*c*) the east–west angular sizes of different brightness temperature regions ($\bullet$, 7,500 K, $\circ$, 10,000 K, $\blacktriangle$, 15,000 K, $\square$, 20,000 K and $\star$, 30,000 K) as a function of time after outburst. The 1σ error is less than $\pm 10\text{ mas}$ for all points.

model in which, as the cooler outer layers of the gas expand and become optically thin, we see the emergence of progressively hotter regions below. On this model the gradients of the curves of Fig. 2b and c are caused by physical expansion together with opacity effects. With this in mind, we calculated a value for the rate of expansion of the ejecta by assuming that the cooler gas (the 7,500 K contours on the maps) had expanded at constant velocity from zero size at the time of outburst until the final epoch. The rate of expansion is equal to 1.5 and 1.1 mas day^{-1} in the north–south and east–west directions respectively. We may derive a distance to Nova Cygni of $1.8 \pm 0.4\text{ kpc}$ from the typical absolute visual magnitude -5.6 for a classical nova 15 days after outburst¹⁴ and a measured apparent visual magnitude of 6.5 (ref. 15), taking a value of $E(B-V) = 0.3$ magnitudes⁴. This would in turn imply a velocity of $2,000 \pm 200\text{ km s}^{-1}$, which is consistent with velocities derived from ultraviolet absorption troughs a few days after outburst¹⁶.

Using data from the most recent epoch, we can estimate limits on electron densities (N_e) of the gas in shells around the nova, assuming an optical depth > 1 (R. M. Hjellming, personal communication). For this condition, the emission measure ($\epsilon = \int N_e^2 dV$) is given by

$$\epsilon = 6.62 \times 10^2 T_b^{1.35} \text{ cm}^{-6} \text{ pc}$$

for an observing frequency of 5 GHz. In each shell we obtain a brightness temperature from the most recent map and integrate

the derived N_e to obtain a number of electrons per shell. We then obtain a mass of protons for each shell, and integrating this gives a total mass for the ejecta of $> 7 \times 10^{-5} M_\odot$. Assuming this mass to be moving at the velocity quoted above, we estimate that the total kinetic energy of the nova is $> 3 \times 10^{38}$ J. $\square$

Received 22 March; accepted 28 April 1993.

1. Sparks, W. M., Starrfield, S. & Truran, J. W. in *Novae and Related Stars*, (ed. Friedjung, M.) 189–204 (Reidel, Dordrecht, 1977).
2. Collins, P. *IAU Circ.* No. 5454 (1992).
3. Payne-Gaposhkin, C. *The Galactic Novae* (North-Holland, Amsterdam, 1957).
4. Shore, S. N. *et al.* *IAU Circ.* No. 5523 (1992).
5. Quirrenbach, A. *IAU Circ.* No. 5463 (1992).
6. Hjellming, R. M. *IAU Circ.* No. 5502 (1992).
7. Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K. & Witzel, A. *Astr. Astrophys.* **61**, 99–106 (1977).
8. Clark, B. G. *Astr. Astrophys.* **89**, 377–378 (1980).
9. Pavellin, P. E., Davis, R. J., Morrison, L. V. & Bode, M. F. *IAU Circ.* No. 5463 (1992).
10. Skiff, B. *IAU Circ.* No. 5461 (1992).
11. Ivison, R., Bang, M. & Bode, M. *IAU Circ.* No. 5516 (1992).
12. Taylor, A. R., Seaquist, E. R., Hollis, J. M. & Pottasch, S. R. *Astr. Astrophys.* **183**, 38–46 (1987).
13. Padin, S., Davis, R. J. & Bode, M. F. *Nature* **315**, 306 (1985).
14. Cohen, J. G. *Astrophys. J.* **292**, 90–103 (1985).
15. Hurst, G. *Astronomy Now* **6**, 41–44 (October 1992).
16. Sonneborn, G., Polidan, R. & Starrfield, S. *IAU Circ.* No. 5456 (1992).
17. Seaquist, E. R. in *Classical Novae* (eds Bode, M. F. & Evans, A.) 143–161 (Wiley, Chichester, 1989).

ACKNOWLEDGEMENTS. P. E. P. thanks the SERC for a Research Studentship.

Narrow lanes of transverse magnetic field in sunspots

H. Zirin & H. Wang

Big Bear Solar Observatory, California Institute of Technology, Pasadena, California 91125, USA

SOLAR flares are closely associated with magnetic activity on the surface of the Sun. They typically occur¹ in complex sunspot groups, where the vertical magnetic fields reverse abruptly, and the horizontal (transverse) fields connecting the vertical poles are both sheared and strong. A single field inversion line may be the site of many flares¹. Here we report observations of a large, active sunspot group which reveal a series of oppositely directed vertical-field inversions separated by extremely narrow elongated channels of intense horizontal fields. In a minimum-energy configuration, lines of force connecting oppositely directed vertical fields simply arch across the inversion line; but when newly emerged sunspots move through older magnetic-field configurations, the poles are pushed together and the field lines turn sharply along the inversion line to reconnect with the vertical field some distance away. These multiple channels of horizontal field imply a large curl ($\nabla \times B$), and hence a substantial electric current. Our observations show that almost all of the larger flares in this region occur in these highly convoluted fields.

We observed the great spot group of June 1991 (NOAA 6659) with the vector videomagnetograph at the Big Bear Solar Observatory. This instrument measures the longitudinal and transverse Zeeman effect by subtracting images in opposite circular or linear polarization. The field of view of the 512×481 pixel frame is 200×150 arcsec ($151,000 \times 105,000$ km). In Figs 1a and 2a we show the relative intensity of transverse and longitudinal fields, irrespective of polarity, on 7 and 10 June. The intensity of each pixel indicates the field strength and the colour indicates the tilt. Red indicates a longitudinal field, directed along the line of sight; green, transverse, in the plane of the sky. Yellow represents a mixture, 45° to the sky.

Sunspots are regions of strong magnetic field consisting of a dark umbra, where the field is vertical, and a fibrous penumbra, where the field lines tilt over and spread out horizontally. A simple round sunspot would be red surrounded by yellow and finally green. The fields around the upper part of spot p in Fig.

2b are an example of such structure, with radiating penumbral fibrils tracing the field lines. As there are no monopoles, the field lines usually arch gracefully to the opposite poles. But if the spot group is complex, spots of opposite polarity may be close together, and sharp boundaries (inversion lines) may occur between oppositely directed field lines. We find that the green inversion lines here may be multiple, with connecting field lines directed along the inversion rather than across it. This is seen in Fig. 1a at inversions 1, 2 and 3, and in Fig. 2a between regions 4 and 7. These are the places where almost all the flares in our region occurred.

Figure 1b shows the longitudinal videomagnetogram, where light and dark pixels indicate positive and negative flux, respectively. NOAA 6659 was highly unusual, showing huge, unbalanced negative magnetic polarity at an unusually high latitude (33° N). Because the opposite-polarity spots P and F

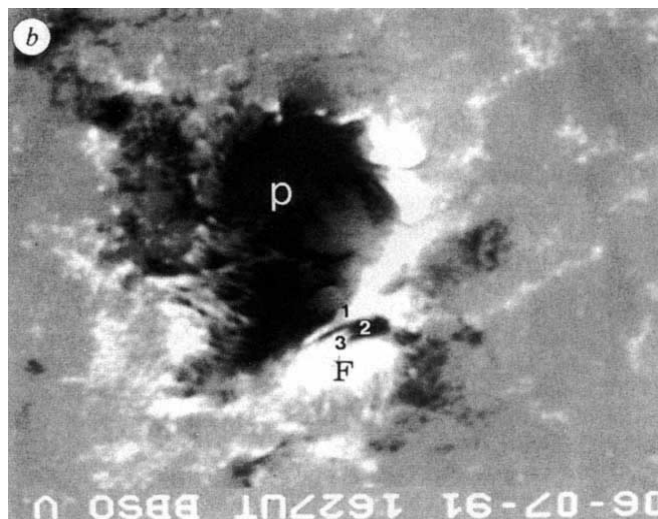
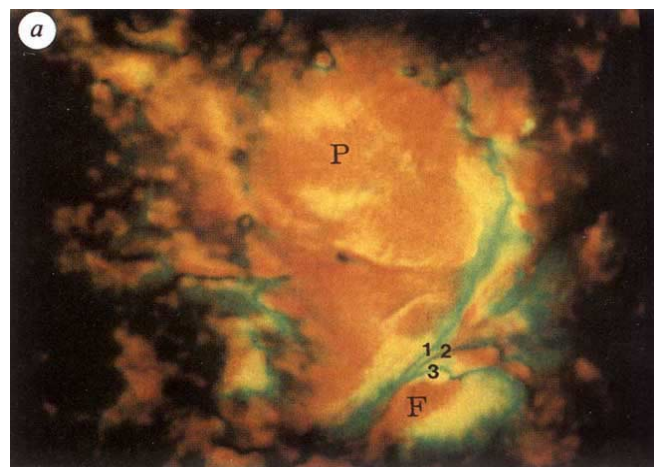


FIG. 1a Colour videomagnetogram taken at a wavelength of $6,103 \text{ \AA}$ of the 7 June 1991 sunspot, showing the tilt of magnetic field lines as determined by the calibrated longitudinal and transverse Zeeman measurements. The colour is determined by the tilt of the field lines: red fields are along the line of sight, yellow at 45° and green transverse to it. The brightness is proportional to the field strength. North is to the top, west to the right on all frames. The narrow green lanes are regions of strong transverse fields separating regions of opposite longitudinal polarity. The spots P and F are opposite polarity, separated by channels marked 1, 2 and 3 of strong transverse field. b, A longitudinal videomagnetogram of same region as a and giving the sign and strength of the fields. White is positive. Here the numbers 1–3 are shifted to mark the alternating longitudinal field regions below each channel. All frames except Fig. 3 are $150,000$ km wide.

were both in the same penumbra, the configuration is classified as a δ -spot, the most active configuration¹. All of the great flares in the region occurred on the boundary between F and P. The numbers 1, 2 and 3 mark the alternating field areas below each corresponding green inversion line in Fig. 1a. In Fig. 3 the transverse field on 7 June is shown (with an ambiguity of 180°); the arrows give the direction and strength of the transverse field and show how it tends to run parallel to the F-P boundary (arrows are omitted from Figs 1a and 2a, where the pixel colour gives the field). The 'half-intensity' width of the green transverse lanes (that is, the separation between points at which lines of force are inclined by 45°) is as small as 3 arcsec, or 2 arcsec when corrected for the 2-arcsec instrument function. Our spectra² show Zeeman splitting corresponding to 3.5 kG in the channel, and the field parallel to it, but they do not resolve the field reversals. All the great flares in the region occurred here. The P-F channel formed before the spot rotated onto the disk, and

was present during the entire disk passage, continuously changing.

Figure 2a is a similar map of the spot on 10 June, near the solar meridian; Fig. 2b shows the corresponding photospheric image in the wing of the emission line at $6,103\text{\AA}$. A series of new spots had emerged overnight; as each dipole expanded, the lines of force were dragged out to form the inversion lines and channels at 4-7. Negative (dark) penumbral fibrils coming out of P were sheared between P and 4. The latter is a region of field lines tilted down 45° in a plane tangent to the umbral boundary and separated by a broad green channel (4A) from the line of spots 5 and 6. The fields in 4A are 1 kG, tilted upwards by about 10° and running along the long axis (vector videomagnetograms are calibrated by comparison to spectra). Areas 5 and 6 have the same polarity as P and are elongated along their direction of motion. A very narrow transverse channel separates these from the larger positive spot 7, with longitudinal field strength 1.5 kG.

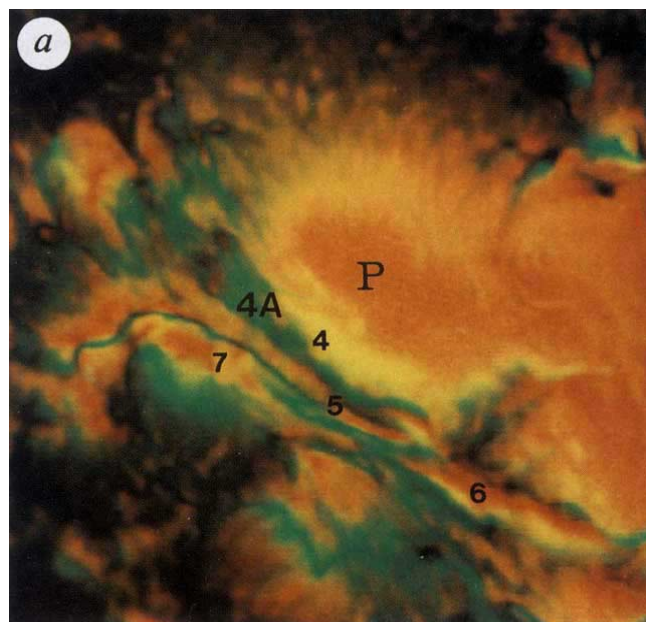
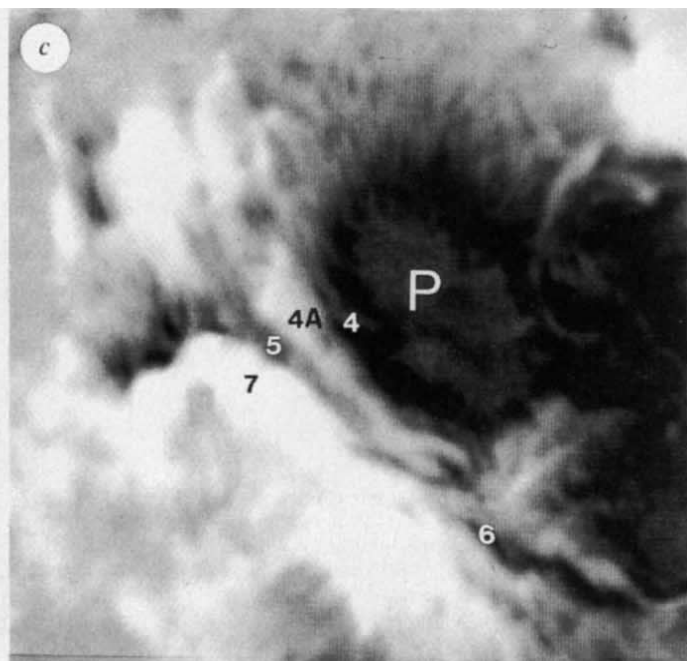
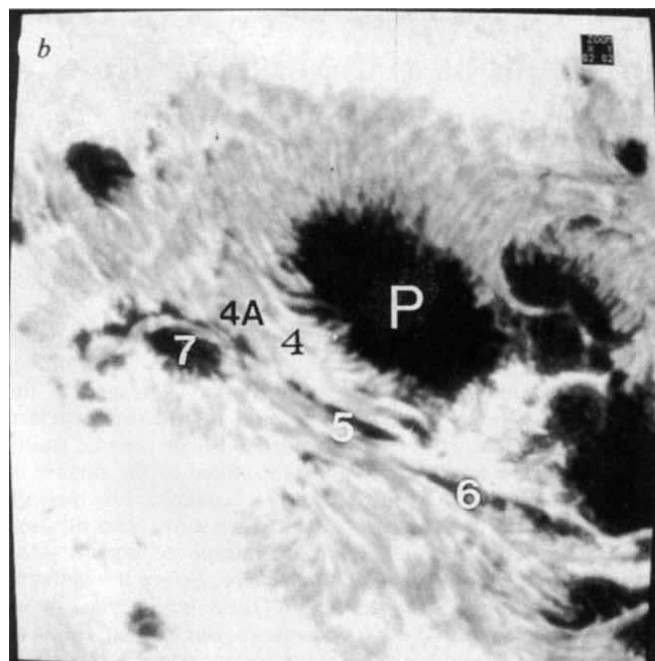


FIG. 2a Magnetogram similar to Fig 1a for the spot on 10 June. Because the spots were spreading, a smaller fraction of the group is visible. New flux emergence at the left edge of P produced a series of elongated umbrae and channels. 4 is a region tilted 45° downwards, parallel to 4A, a region of transverse penumbral fibrils; 5 and 6 are two elongated umbrae of the same polarity as P, and 7 is a new spot of F (white) polarity. b, CCD image of the spot with a broad-band filter near $8,500\text{\AA}$, showing penumbral structure. Corresponding numbers are marked. c, Longitudinal videomagnetogram (like Fig. 1b) of the spot, with regions marked.



There were many flares here, all smaller than those near F, where fields were twice as strong. The channel structure formed in half a day and lasted for several days.

These magnetograms show the field at the intersection of a three-dimensional cluster of bipolar loops with the surface of optical depth unity (the photosphere), where the spectrum line measured originates. When we image the Sun in a much stronger line, such as $H\alpha$, we see the overlying chromosphere, where flares occur in magnetic configurations determined by the photospheric fields. The $H\alpha$ structure shows fibrils parallel to the transverse fields; by matching these to the polarities in the magnetograms, we can remove the 180° ambiguity in transverse field direction. Our notion of the resulting three-dimensional field structure is sketched in Fig. 4, which shows the field line configuration generated by a moving spot. Each dipole is a horizontal slice of a flux loop, so as these loops break the surface, the poles separate at about 1 km s^{-1} . Because the field lines are stuck to the original poles, they are elongated by the motion and a sharp turn at the neutral (or inversion) line results.

If all flux loops rose at the same rate, there would only be self-similar growth and no relative motion or stress or shear on the field lines. More often a main spot emerges and new dipoles emerge later at different rates. Each flux loop expands into the older fields, but also reconnects with them³, and field lines stretch along the expansion track of the new spots. The flux loops are thought to rise because of magnetic buoyancy, the tendency of enclosed flux loops to expand outward and rise by buoyancy^{4,5}, which pushes new field lines up through old. Such developments are typical of the most frequent and intense activity. Each additional channel represents a corresponding increase in the strong field-aligned currents which give rise to flares. When flux emergence ceases, the activity dies out.

These observations provide unequivocal evidence of narrow lanes of transverse field in flaring regions. Hints of such lanes appear in direct images and longitudinal magnetograms taken in 1988⁵. We are sure that the present high-resolution vector magnetograph would have shown similar field structure in those cases.

Because the field energy is the volume integral of B^2 , the

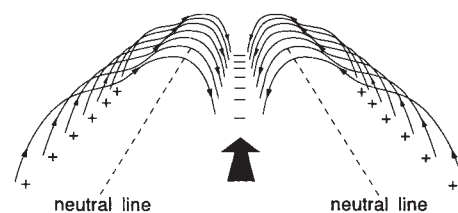


FIG. 4 Schematic diagram of the field line configuration produced by an elongated spot of negative polarity moving (arrow) between two positive spots, such as area 5 in Fig. 2a. The lines are vertical over the spots (red in 2a), then turn sharply parallel to the neutral line (green channel), and dive back to the surface at the next spots (red). In the absence of a current, the field lines would simply arch across the channel to the nearest opposite polarity.

appearance of several channels in a single boundary does not affect the energy, but each reversal produces an additional current. If the field reverses in $1,000 \text{ km}$, as observed, the field-aligned current density ($\nabla \times \mathbf{B}$) along a cylindrical channel would be 0.5 A m^{-2} , an order of magnitude larger than derived from lower-resolution observations⁶. The total current of each channel is $5 \times 10^{11} \text{ A}$.

Because solar flares occur where magnetic shear (defined as the difference between observed and potential field configurations) is large, we assume that they release magnetic energy. Paradoxically, recent data for five flares^{7,8} show that shear increases. However, indirect methods^{9,10} suggest that shear decreases, which one might expect intuitively. Whatever really happens, we have shown that the flare-related magnetic fields are far more complex and convoluted than previously thought. □

Received 5 October 1992; accepted 20 April 1993.

1. Zirin, H. *Astrophysics of the Sun* (Cambridge Univ. Press, 1988).
2. Zirin, H. & Wang, H. *Solar Phys.* **144**, 37 (1993).
3. Zirin, H. & Wang, H. *Solar Phys.* **125**, 45 (1990).
4. Parker, E. N., *Astrophys. J.* **121**, 491 (1955).
5. Zirin, H. *Astrophys. J.* **274**, 900 (1983).
6. de La Beaujardiere, J. F., Canfield, R. C. & Leka, K. D. *Astrophys. J.* (submitted).
7. Wang, H. *Solar Phys.* **140**, 85 (1992).
8. Wang, H., Ewell, M. W., Zirin, H. & Ai, G. *Astrophys. J.* (submitted).
9. Sakurai, T. et al. *Publ. Astr. Soc. Japan* **44**, L123 (1992).
10. Neidig, D. *Solar Phys.* **61**, 121 (1979).

ACKNOWLEDGEMENTS. We thank G. Eychaner and J. Varsik for their work on the vector-magnetograph program, and M. W. Ewell for reading the manuscript. This work was supported by the NSF, NASA and ONR. The vector magnetograph was built with a grant from Caltech.

Implications of the high D/H ratio for the sources of water in Venus' atmosphere

David Harry Grinspoon

Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA

THE high abundance ratio of deuterium to hydrogen in the atmosphere of Venus (120 times that on Earth) can be interpreted either as the signature of a lost primordial ocean¹, or of a steady state in which water is continuously supplied to the surface of Venus by comets or volcanic outgassing, balancing loss through hydrogen escape^{2,3}. New observations⁴⁻⁶ of a water concentration of only 30 parts per million in Venus' atmosphere imply that the residence time of water in the atmosphere, before it escapes to space, is short compared with the age of the Solar System, casting doubt on the primordial ocean hypothesis. But a recent theoretical reanalysis of collisional ejection⁷ has increased estimates of

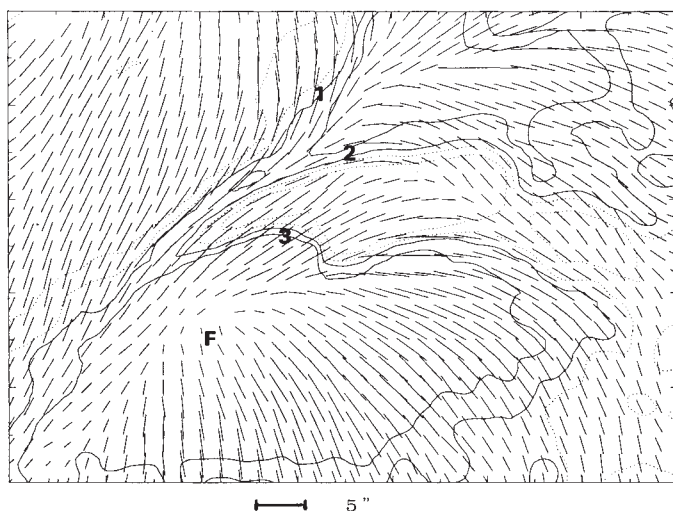


FIG. 3 Magnified vector magnetogram corresponding to Fig. 1a, showing the channel structure 1-2-3 on 7 June. Solid contours are for positive, dotted for negative longitudinal field. The bars give the transverse field direction ($\pm 180^\circ$) and strength for 3×3 pixels at their centre. Except for the left part of channel 3, the bars run parallel to the channels. We find this alignment in virtually all δ spots. The apparent strong transverse field at lower right is the limbward projection of the longitudinal field.

the deuterium escape efficiency by a factor of 10: this means that if the venusian water budget is in steady state, the D/H ratio of the source water must be 10–15 times higher than that on Earth, ruling out cometary water, whose D/H ratio is thought to be lower than this⁸. Here I suggest that these observations can be understood either as the result of continuous outgassing from a highly fractionated mantle source (such as might result from severe dessication of the mantle, or massive hydrogen escape early in the planet's history) or Rayleigh fractionation after massive outgassing from catastrophic resurfacing of the planet in the past 0.5–1 Gyr.

Estimates of the total escape flux of hydrogen from Venus cover a range from 0.4×10^7 to $3.7 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$ (refs 9–14). These fluxes are highly uncertain as they depend on the temperature structure of the nightside ionosphere and on the response of the atmosphere to changes in the solar wind, both of which are poorly known. Because the combined uncertainties in these estimates are greater than the differences among them, and because they are likely to change further as theory and observations advance, I will use this range of estimated fluxes as an indicator of the uncertainty and take a median escape flux of $2 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$ as a nominal value for evolutionary modelling.

Several processes contribute to hydrogen escape, each with a different efficiency of deuterium escape relative to hydrogen, represented by the fractionation factor

$$f = \frac{(dD/dr)/D}{(dH/dr)/H} \quad (1)$$

where D and H are the atmospheric abundances of deuterium and hydrogen. Recently Gurwell and Yung⁷ have reanalysed the collisional ejection of H and D. This process, the second most important escape process for H, was previously thought to be extremely inefficient at allowing D to escape, but is now found to be its main escape mechanism. (For H, the most important process is charge exchange with high-temperature ions, then collisions with fast O atoms then direct ion escape in the plasma tail.) They find a weighted fractionation factor, averaged over the three important escape processes, of $f = 0.13$, an order of magnitude higher than previous estimates.

A new spectroscopic analysis of the near-infrared windows on the nightside of Venus yields a water abundance of 30 ± 10 p.p.m. (refs 4, 5). Previous spacecraft measurements had left an order-of-magnitude uncertainty in the water abundance, estimated at 20–200 p.p.m. (refs 1, 14–16). The Pioneer Venus large-probe neutral mass spectrometer data, originally interpreted to give a substantially greater water abundance¹⁶, have been re-analysed⁶ to yield 28 p.p.m., in agreement with the new spectroscopic determinations. A consensus thus seems to have emerged on the present abundance of water on Venus. This remarkably desiccated atmosphere has a lifetime against H escape of 74 million to 680 million years, for the escape fluxes given above. Even the high end of this range is much shorter than the age of the Solar System. This suggests that water on Venus now is probably not the remnant of a large ancient reservoir. With these relatively short lifetimes, more recent sources are probable and a steady state is more likely.

The time-dependent solution for evolution of D/H in an atmosphere with a steady-state water abundance is given by^{2,17}

$$\left(\frac{D}{H}\right)(t) = \frac{\alpha}{f} - \left[\frac{\alpha}{f} - \left(\frac{D}{H}\right)_0\right] \exp(-\phi ft/H_{ss}) \quad (2)$$

where H_{ss} is the steady-state hydrogen column abundance, α is the D/H ratio in the source, ϕ is the flux of H through the system in $\text{cm}^{-2} \text{ s}^{-1}$, and $(D/H)_0$ is the D/H ratio at time $t = 0$. Here the time-dependent term represents the signature of the original D/H ratio which decays asymptotically with time constant

$$\tau_{ss} = \frac{H_{ss}}{\phi f} = \tau_{H,0}/f \quad (3)$$

After a steady state has existed for several times τ_{ss} , the D/H

ratio decays towards the steady-state limit, $(D/H)_{lim} = \alpha/f$. I will refer to this limiting state as a 'mature steady state'. This time constant is longer than the lifetime of water by a factor of $1/f$, as indicated by equation (3), so that water reaches steady state well before the D/H ratio reaches the limiting value. The behaviour of equation (2) for a range of parameters is shown in Fig. 1. Only after several times τ_{ss} does the system achieve a mature steady state, with the D/H ratio relaxing towards $(D/H)_{lim}$.

The theoretical work described above raises the weighted fractionation factor by an order of magnitude over the value used in earlier evolutionary modelling^{2,14}. As both $(D/H)_{lim}$ and τ_{ss} vary inversely with f , this change has a paradoxical effect on the interpretation of the significance of the observed D/H ratio for the history of water on Venus. Assuming that the source hydrogen has a terrestrial D/H value ($\alpha = 1.6 \times 10^{-4}$), then the higher value of f gives $(D/H)_{lim} = 1.23 \times 10^{-3}$, well below the observed value for the steady-state limit D/H ratio. The nominal value of τ_{ss} is now reduced to 1.1×10^9 years. Thus we have a paradox: the timescale is now more favourable for the system to have reached $(D/H)_{lim}$ in several billion years ($\tau_{ss} < \text{age of}$

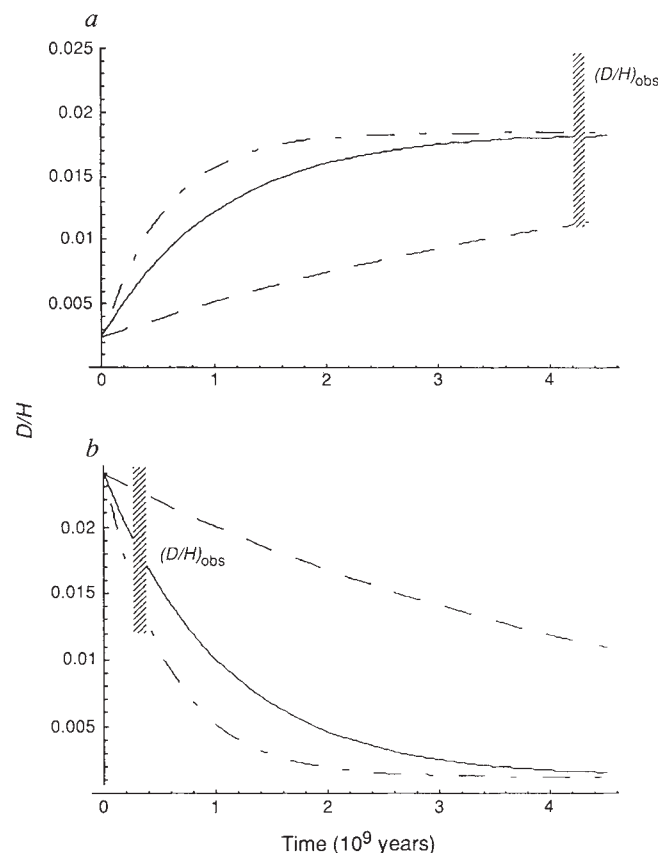


FIG. 1 Hypothetical curves of the evolution of the deuterium/hydrogen ratio with water in steady state, from equation (2) using $f = 0.13$. Solid, dashed and dot-dash curves are for H escape fluxes of 2.0 , 0.4 and $3.7 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$, respectively. These curves show that in steady state, the ratio will eventually converge towards $(D/H)_{lim}$, with the timescale depending on the H flux. Observed ratio shown is 120 ± 40 times the terrestrial value²⁸. *a*, Curves for $\alpha = (D/H)_0 = 15$ times terrestrial, showing that for a H source with an enhancement of this order, steady state can produce the observed D/H. *b*, $\alpha = \text{terrestrial}$, and $(D/H)_0 = 150$ times terrestrial, the value calculated in the text for fractionation following catastrophic resurfacing. This curve shows that an enhanced D/H created by Rayleigh fractionation from massive escape would be preserved if this episode occurred in the last 10^9 years, but might well be erased if the episode of fractionating escape was 'primordial'.

Venus), yet the limit has now decreased to well below the observed value ($(D/H)_{\text{lim}} < (D/H)_{\text{obs}}$).

Possible explanations for this dilemma fall into two categories, each with implications for the evolution of Venus. One possibility is that water on Venus is in steady state, with some source (such as volcanic outgassing) balancing hydrogen loss to space. In this case the D/H ratio of the source hydrogen (α) must be substantially higher than that of terrestrial water. A value of α enhanced by a factor of 10 or 15 over the terrestrial value would result in $(D/H)_{\text{lim}} \approx (D/H)_{\text{obs}}$ (Fig. 1a). This would preclude comets from being the dominant source of water if the observations of D/H for comet Halley⁸ are representative.

If the atmosphere is in steady state with an outgassing source, then the H escape flux provides a direct measure of the time-averaged outgassing flux of hydrogen-containing gases over the last $\tau_{\text{H}_2\text{O}} \approx 130$ million years. An H escape flux of $\sim 2.0 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$ gives an upper limit for the outgassing flux of $4.3 \times 10^{10} \text{ g}$ of water per year. This estimate may be combined with recent observations by the Magellan spacecraft to evaluate the nature of outgassing and the likelihood of steady state with a fractionated mantle source.

Magellan mapping of volcanic features¹⁸, combined with simulations of the crater population^{19–21}, has allowed preliminary estimates for volcanic resurfacing rates of $\sim 0.4 \text{ km}^3 \text{ yr}^{-1}$ (ref. 21), yielding a rate of magma production of $\sim 1.2 \times 10^{15} \text{ g yr}^{-1}$. Combining this with the outgassing flux yields a lower limit for the average water content of venusian magmas of ~ 50 p.p.m. by weight, assuming efficient outgassing of lava flows. This is more than 1,000 times drier than typical basaltic magmas on Earth²². Even if outgassing is inefficient, venusian magmas are apparently extremely desiccated, perhaps implying a desiccated mantle. A desiccated mantle could be the result of a history of outgassing with no recycling of water in the absence of a surface reservoir and only limited subduction. Could this mantle reservoir be the source of the fractionated hydrogen necessary to reconcile $(D/H)_{\text{obs}}$ with $(D/H)_{\text{lim}}$? The D/H fractionation factor for mantle degassing has not been modelled, and is probably near unity. Large fractionation effects are found, however, in severely depleted reservoirs. If water in the mantle has been depleted by a factor of 10^5 (roughly the difference between the known water reservoirs of Venus and Earth), then an outgassing fractionation factor of $f = 0.8$ could have enhanced D/H in the mantle by a factor of 10. Alternatively, an enhanced mantle ratio could reflect massive hydrogen escape which occurred very early in the planet's history and was frozen into the mantle when an early magma ocean solidified. If so, the observed D/H may be the result of primordial water loss, but this signature would have been preserved in the mantle, not in the atmosphere.

Another possibility is that there has been a massive injection of unfractionated water within the past 0.5–1 billion years followed by massive fractionating escape. If this occurred less than several times τ_{ss} ago, then the enhanced D/H from this episode would be preserved whether or not venusian water is currently in steady state. It may be relevant that the cratering record of Venus (as revealed by Magellan) is peculiar in that the surface is young (compared with other non-terrestrial surfaces in the inner Solar System) and yet there is little evidence of any ongoing processes which are rapidly removing craters. This has led to the hypothesis that a catastrophic resurfacing occurred 0.5–1 Gyr ago, covering up all earlier craters²⁰. This idea, still controversial, has received support from Monte Carlo models of surface evolution²¹, and from geochemical and thermal models^{23,24} which suggest that the interior may go through periods of episodic overturn with a frequency of ~ 0.5 Gyr. Alternatively, the interior and surface may simply have been continuously more active before that time.

Much higher resurfacing rates in the relatively recent past would have important consequences for atmospheric and climatic evolution. The volcanic gases SO_2 and H_2O are the chief

contributors to the strong atmospheric greenhouse²⁵. Catastrophic resurfacing would greatly increase the outgassing rates and thus the abundances of these gases, until escape processes, photochemistry and surface reactions returned abundances to their equilibrium values. The volume of magma necessary to resurface the planet deeply enough to cover all pre-existing craters is approximately $5 \times 10^9 \text{ km}^3$ (ref. 23). Scaling the current estimated resurfacing rate and hydrogen escape flux, this episode, which is estimated to have occurred over a period of 10^7 – 10^8 years²³, could have injected up to ~ 150 times the current water inventory, or 4,500 p.p.m. water, into the atmosphere. The fractionation factor and timescale for the ensuing rapid hydrogen escape would depend on which escape processes dominated. Diffusion-limited escape could remove this much water on a timescale of several hundred million years, but the expected escape processes and rates for a venusian atmosphere with greatly increased water abundances are only partially understood^{26,27}. The isotopic effects of this rapid escape are best modelled as a Rayleigh fractionation^{1,2}

$$\left(\frac{D}{H}\right)_1 = \left(\frac{D}{H}\right)_0 \left[(X_{\text{H}_2\text{O}})_0 / (X_{\text{H}_2\text{O}})_1\right]^{1-f} \quad (4)$$

where $X_{\text{H}_2\text{O}}$ is the mixing ratio of water, is assumed to represent the total hydrogen inventory and the subscripts 1 and 0 indicate the values at present, and at the start of the fractionating escape, respectively. Escape of 150 times the current water inventory, with the fractionation factor of 0.13 estimated for the current atmosphere, would enhance the D/H ratio by a factor of ~ 80 . The uncertainty in this estimate is at least a factor of several hundred per cent, but it demonstrates that a recent massive outgassing event (or very massive comet impact(s)) could cause an enhancement of the D/H ratio of the order observed. Thus $(D/H)_{\text{obs}}$ may be the isotopic signature of catastrophic resurfacing in the past 0.5–1 Gyr.

This interpretation requires no current hydrogen source. After the massive injection of water, escape processes would decrease the water abundance, with a concurrent increase in D/H due to Rayleigh fractionation, until the escape flux declined to a rate equal to the hydrogen source flux (equal to the time-averaged flux of cometary hydrogen plus any outgassing contribution). It is not possible with current observational uncertainties to determine whether water is at present declining or in steady state. After the system returned to steady state, the enhanced D/H ratio would decline asymptotically towards $(D/H)_{\text{lim}}$, as seen in Fig. 1b. This figure demonstrates that if a steady state exists it could not have greatly affected the D/H ratio in the time since the hypothesized catastrophic resurfacing. Figure 1b also shows that $(D/H)_{\text{obs}}$ is probably not the signature of primordial water escape, which would have disappeared by now. In contrast, the timescale for the hypothesized catastrophic resurfacing is consistent with τ_{ss} .

Alternatively, if the actual hydrogen escape flux is lower than current theory suggests, τ_{ss} could be sufficiently long that the atmosphere would not approach $(D/H)_{\text{lim}}$ over the age of the planet. In this case $(D/H)_{\text{obs}}$ could be the remnant of a period of massive fractionating escape early in the planet's history, yet the relationship between $(D/H)_{\text{obs}}$ and the 'primordial' water inventory would be difficult to determine. The extremely dry atmosphere seems to preclude this possibility. The mass of water corresponding to a mixing ratio of 30 p.p.m. is roughly 10 times the water content of comet Halley. Even without any outgassing source over the history of the planet, primordial water may have been replaced many times over by comet impacts and subsequent escape. For more certain assessment of the alternatives, we need more detailed models of escape mechanisms and the response of the atmosphere to large water abundances. □

Received 1 September 1992; accepted 8 April 1993.

1. Donahue, T. M., Hoffman, J. H., Hodges, R. R. & Watson, A. J. *Science* **216**, 630–633 (1982).

2. Grinspoon, D. H. *Science* **238**, 1702–1704 (1987).
3. Grinspoon, D. H. & Lewis, J. S. *Icarus* **74**, 21–35 (1988).
4. Pollack, J. B. *et al.* *Icarus* (in the press).
5. de Bergh, C. *et al.* *Adv. Space Res.* (in the press).
6. Donahue, T. M. & Hodges, R. R. *Papers presented to int. Colloq. Venus, LPI Contrib. No. 789* (1992).
7. Gurwell, M. A. & Yung, Y. L. *Planet. Space Sci.* (in the press).
8. Eberhardt, P. *et al.* *Astr. Astrophys.* **187**, 435–437 (1987).
9. Rodriguez, J. M., Prather, M. J. & McElroy, M. B. *Planet. Space Sci.* **32**, 235–235 (1984).
10. Donahue, T. M. & Hartle, R. E. *Geophys. Res. Lett.* **19**, 2449–2452 (1992).
11. Brace, L. H. *et al.* *J. geophys. Res.* **92**, 15–26 (1987).
12. Hodges, R. R. & Tinsley, B. A. *J. geophys. Res.* **91**, 13649–13658 (1986).
13. Grinspoon, D. H. *Papers presented to int. Colloq. Venus, LPI Contrib. No. 789* (1992).
14. Hunten, D. M., Donahue, T. M., Walker, J. C. G. & Kasting, J. F. in *Origin and Evolution of Planetary and Satellite Atmospheres* (eds Atreya, S., Pollack, J. B. & Matthews, M. S.) 386–422 (Univ. of Arizona Press, 1989).
15. Lewis, J. S. & Grinspoon, D. H. *Science* **249**, 1273–1275 (1990).
16. Donahue, T. M. & Hodges, R. R. *J. geophys. Res.* **97**, 6083–6091 (1992b).
17. Krasnopolsky, V. A. *Icarus* **62**, 221 (1985).
18. Head, J. W., Crumpler, L., Aubele, J., Guest, J. & Saunders, R. S. *J. geophys. Res.* **97**, 13153–13197 (1992).
19. Phillips, R. J. *et al.* *J. geophys. Res.* **97**, 15923–15948 (1992).
20. Schaber, G. G. *et al.* *J. geophys. Res.* **97**, 13257–13302 (1992).
21. Bullock, M. A., Grinspoon, D. H. & Head, J. W. *Papers presented to int. Colloq. Venus, LPI Contrib. No. 789* (1992).
22. Moore, J. G. *Contrib. Mineral. Petrol.* **28**, 272 (1970).
23. Head, J. W., Parmentier, E. M. & Hess, P. C. *Papers presented to int. Colloq. Venus, LPI Contrib. No. 789* (1992).
24. Turcotte, D. *Papers presented to int. Colloq. Venus, LPI Contrib. No. 789* (1992).
25. Pollack, J. B., Toon, O. B. & Boese, R. J. *geophys. Res.* **85**, 8223–8231 (1980).
26. Kumar, S., Hunten, D. M. & Pollack, J. B. *Icarus* **55**, 369–389 (1983).
27. Kasting, J. F. & Pollack, J. B. *Icarus* **53**, 479–508 (1983).
28. de Bergh, C. *et al.* *Science* **251**, 547–549 (1991).

ACKNOWLEDGMENTS. I thank M. Bullock, T. Colaprete, J. Holloway, J. Kasting, J. Pollack and J. Head for discussions and K. Zahnle for a thorough review.

Nanometre-scale recording and erasing with the scanning tunnelling microscope

Akinobu Sato & Yuji Tsukamoto

Functional Devices Research Laboratories, NEC Corporation, 4-1-1, Miyazaki, Miyamae-ku, Kawasaki, Kanagawa 216, Japan

DATA storage on magnetic and optical disks and in semiconductor devices is being achieved at ever higher areal densities^{1,2}. The use of the scanning tunnelling microscope (STM) for both data storage and nanofabrication has been explored recently^{3–9}. These approaches use the STM tip to produce nanometre-scale marks or structures on a surface, sometimes with atomic precision. Most studies have, however, neglected erasure of recorded marks (an exception is described in ref. 10), despite the fact that erasure is essential for any practical recording device. Here we demonstrate the use of the STM for reproducible and reversible recording and erasing of marks about 10 nm in size. These are written at ambient temperature and pressure onto the surface of a composite medium consisting of a thin layer of a vanadate glass deposited on vanadium bronze, β - $\text{Na}_x\text{V}_2\text{O}_5$. The present recording speed of 1 ms is several orders of magnitude too slow for practical applications; this remains a challenge for future study.

Using single atoms as the memory units in a data storage device is an ultimate goal of attempts to manipulate atoms with the STM and atomic force microscope (AFM). But manipulation of atoms requires ultra-high vacuum, low temperatures and long times³. Moreover, reversibility is one of the most important factors for data storage devices. In those approaches that involve changes in surface topography — for example, where the local tip pressure of the AFM creates a mark⁵ — the mark is impossible to erase. In addition, it becomes hard to distinguish between recording marks and pre-existing surface defects, especially in smaller marks. Nanometre-scale chemical modification, using STM and compounds of composition Ag_2Se , has

been reported⁶, but again erasure is prohibited because the selenium atoms are removed from the surface of the medium by reaction with ambient hydrogen.

Basing recording and reproduction on changes in tunnelling current without any change in surface morphology seems a more suitable way to produce an STM memory. It is well known that phase transitions, for example from amorphous to crystalline, change electrical properties. If a phase change can be induced locally and in a controlled manner by using an STM, it might be possible to achieve all of the necessary storage functions by modifying the tunnelling current.

Chalcogen compounds and vanadate glasses, which show a phase transition accompanied by electrical conductivity changes, were examined as possible recording media. Both materials proved inadequate, however, because of chemical instability for the former and irreversibility for the latter. We finally adopted as a potential recording medium a composite material comprising a crystalline vanadate and its glassy, amorphous form¹¹.

The glass was prepared by rapid quenching¹¹. The glass composition obtained was 14 mol% Na_2O –71 mol% V_2O_5 –15 mol% P_2O_5 . The V_2O_5 is an electrical conduction component. Crystal phase stability and the mechanism of electrical conduction depend on the Na_2O concentration. The P_2O_5 was added to stabilize the amorphous form. The glass was heat-treated at 573 K for 1 hour to crystallize it. Deposited crystals were identified as β - $\text{Na}_x\text{V}_2\text{O}_5$, $x \approx 0.33$ ('vanadium bronze') by X-ray diffraction. The vanadium bronze possesses unique electrical properties, including quasi-one-dimensional conduction and small polaron behaviour^{12,13}. The amorphous phase in the crystallized glass can be classified into two states. One is a heterogeneous phase, more than 100 nm thick. The other is a uniform layer less than 10 nm thick on the *bc*-plane of the vanadium bronze crystal. The latter was suitable for a reversible recording and erasing medium.

Figure 1A shows a $100 \times 100 \text{ nm}^2$ top-view STM image after four marks have been recorded. The bias voltage for imaging was +0.7 V. The mark size was about $10 \times 10 \text{ nm}^2$, corresponding to $\sim 1 \times 10^{12}$ bits per square inch area density. This recording was made on 10-nm-thick amorphous material on the *bc*-plane of β - $\text{Na}_x\text{V}_2\text{O}_5$. The marks were formed by applying pulsed voltages of +4 V (the bias polarity refers to the sample) for times of (a) 10 ms, (b) 5 ms, (c) 2 ms and (d) 1 ms. The instantaneous tunnelling current during a pulse of +4 V lasting for 1 ms is about 50 nA. For data storage a recording speed of 1 ms is orders of magnitude too slow. We hope that higher recording speeds might be attained by using as the recording medium materials in which the ion mobility is greater. These can range over several orders of magnitude; for example, the ion conductivity for superionic conducting glass, such as the AgI – Ag_2O – P_2O_5 system¹⁴, can be more than four orders of magnitude higher than for the vanadate glass.

Marks were erased by applying a reverse-polarity pulse of –5 V for 10 ms, as shown in Fig. 1B. This reversible transforma-

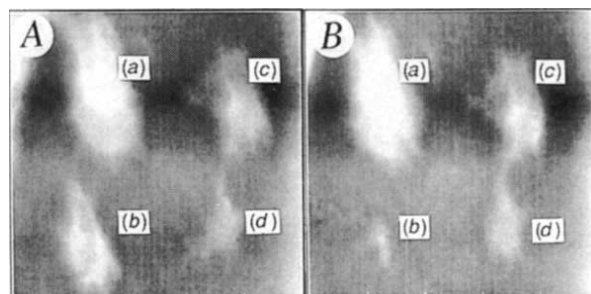


FIG. 1 A, Top-view STM image after recording four marks. The marks were recorded using pulses of 4 V for (a) 10 ms, (b) 5 ms, (c) 2 ms and (d) 1 ms. B, The STM image after applying a reverse-polarity voltage pulse (–5 V for 10 ms) to mark (b).

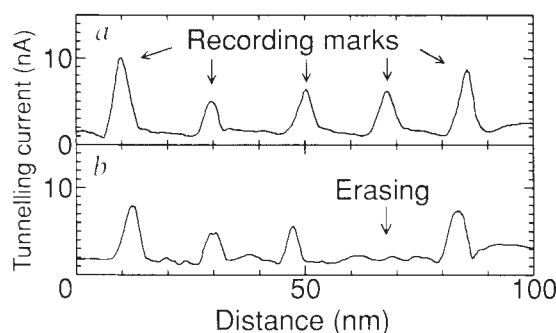


FIG. 2 Tunnelling current waveforms after recording and erasing. *a*, Five marks recorded by application of a 4 V pulse for 1 ms. *b*, The fourth mark shown in *a* has been erased by application of a -5 V pulse for 10 ms.

tion was repeated more than 10 times. Any recording mark subjected to writing and erasing pulses more than 10 times became impossible to erase. The reproduced signal is the change in tunnelling current measured with the STM in constant height mode. Figure 2*A* and 2*B* shows reproduced waveforms after recording five marks and subsequently erasing one mark, respectively.

Figure 3 shows current-voltage (*I-V*) curves, both before and after recording. Before the recording, the *I-V* curve shows characteristics typical of a semiconductor with bandgap (Fig. 3*A*). After recording, the *I-V* curve shows metallic characteristics (Fig. 3*B*). The semiconducting *I-V* and the metallic *I-V* curves were also consistent with those expected for the amorphous phase and crystalline phase in crystallized glasses, respectively. A regular configuration of atoms, corresponding to β - $\text{Na}_x\text{V}_2\text{O}_5$ atomic arrangements, was also observed in the recording mark. We attribute the recording process to a phase transition from the amorphous to the crystalline phase. Deposition of contaminants such as tip material might instead be responsible, but this mechanism would then be expected to hold regardless of the recording medium. We tested both β - $\text{Na}_x\text{V}_2\text{O}_5$ crystals and sputtered gold film by applying a similar pulse voltage to that used in Fig. 1*A*, but no recording mark was observed on either medium. This implies that the recording mechanism for $\text{Na}_2\text{O-V}_2\text{O}_5\text{-P}_2\text{O}_5$ is unlikely to involve contamination.

The reversible recording mechanism can be explained on the basis of sodium ion migration. From previous studies^{11-13,15,16}, it is known that ionized alkali metals have high mobility in both vanadium bronze crystals and vanadate glasses, and that the stabilities of the crystalline and amorphous phases are dependent on sodium concentration. As shown in Fig. 3, the *I-V* characteristics for marks can be changed reversibly by altering the polarity of the applied voltage pulse. Further evidence for sodium ion migration comes from experiments using crystallized

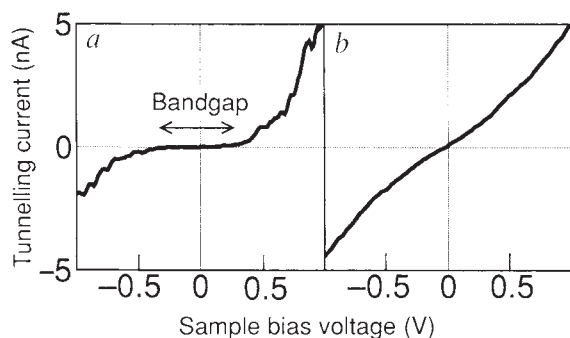


FIG. 3 *I-V* curves (*a*) before recording and (*b*) after recording.

$\text{V}_2\text{O}_5\text{-P}_2\text{O}_5$ glass, which was free from sodium ions and was prepared by the same process as the $\text{Na}_2\text{O-V}_2\text{O}_5\text{-P}_2\text{O}_5$ medium. Recording marks could be made on the $\text{V}_2\text{O}_5\text{-P}_2\text{O}_5$ medium by applying voltage pulses of 5 V for 100 ms. But the marks could not be erased in this medium under any of the pulse voltage conditions tried, including reversing the pulse polarity. These facts suggest that the main mechanism for reversible recording and erasing is sodium ion migration along the electric field. □

Received 13 November 1992; accepted 23 April 1993.

1. Tsang, C., Chen, M., Yogi, T. & Ju, K. *IEEE Trans. Magn.* **26**, 1689-1693 (1990).
2. Futamoto, M. et al. *IEEE Trans. Magn.* **27**, 5280-5285 (1991).
3. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524-526 (1990).
4. Ehrichs, E. E., Smith, W. F. & de Lozanne, A. L. *J. Vac. Sci. Technol.* **B9**, 1380-1383 (1991).
5. Mamin, H. J. & Rugar, D. *Appl. Phys. Lett.* **61**, 1003-1005 (1992).
6. Utsugi, Y. *Nature* **347**, 747-749 (1990).
7. Rabe, J. P. & Buchholz, S. *Appl. Phys. Lett.* **58**, 702-704 (1991).
8. Mamin, H. J., Chiang, S., Birk, H., Guethner, P. H. & Rugar, D. *J. Vac. Sci. Technol.* **B9**, 1398-1402 (1991).
9. Schimmel, Th., Fuchs, H., Akari, S. & Dransfeld, K. *Appl. Phys. Lett.* **58**, 1039-1041 (1991).
10. Rubin, K. A. *J. Magn. Soc. Jap.* **15**(S1), 127-132 (1991).
11. Hirashima, H., Sato, A. & Yoshida, T. *J. Ceram. Soc. Jap.* **94**, 875-879 (1986).
12. Kapustkin, V. K., Volkov, V. L. & Fotiev, A. A. *J. Solid State Chem.* **19**, 359-363 (1976).
13. Goodenough, J. B. *J. Solid State Chem.* **1**, 349-358 (1970).
14. Minami, T., Takuma, Y. & Tanaka, M. *J. Electrochem. Soc.* **124**, 1659-1662 (1977).
15. Scholtens, B. B. *Mat. Res. Bull.* **11**, 1533-1538 (1976).
16. Murawski, L., Gledel, C., Sanchez, C., Livage, J. & Audieres, J. P. *J. Non-Cryst. Solids* **89**, 98-106 (1987).

An atom probe for three-dimensional tomography

D. Blavette, A. Bostel, J. M. Sarrau, B. Deconihout & A. Menand

Laboratoire de Microscopie Ionique, URA CNRS 808, Faculté des Sciences et Techniques BP 118, 76134 Mont Saint Aignan Cedex, France

ELECTRIC-field-induced evaporation of ions from a needle-like surface, and their subsequent identification by time-of-flight mass spectrometry, forms the basis of the atom-probe technique¹. This has proved to be a powerful analytical tool^{2,3}, permitting the quantitative determination of material composition in a small selected region of the surface (depths of 1-5 nm) with single-layer resolution. Positional information for the atoms within each layer is lost, however. In contrast, the field-ion microscope³ provides atomic-resolution images of surfaces, but without information about the nature of the atoms. Several attempts have been made to combine these two techniques by extending the time-of-flight measurement into two dimensions, but they have been limited by their inability to identify all chemical species⁴ or to combine spatial and temporal information for multiple events⁵, especially for ions with very similar mass-to-charge ratios⁶. Here we make use of a recently developed⁷ multiple-impact detector to construct a position-sensitive atom probe with sufficiently high temporal resolution (of the order of 10 ns) to avoid these earlier problems; thus, reliable composition and position data can be obtained at the same time. We illustrate the performance of this instrument by imaging the three-dimensional distribution of chemical heterogeneities in a nickel-based alloy on a near-atomic scale.

Our tomographic atom-probe is depicted schematically in Fig. 1. The sample, prepared in the form of a very sharp tip (radius $R \approx 10\text{-}50\text{ nm}$), is subjected to a constant high positive voltage ($V_0 \approx 5\text{-}20\text{ kV}$). On top of this, high voltage pulses ($V_p \approx 1\text{-}5\text{ kV}$, rise time $< 1\text{ ns}$, duration $\sim 10\text{ ns}$) are applied to the tip; these cause surface atoms to be evaporated and ionized. High vacuum conditions (10^{-8} Pa) are used. The ions originating from a

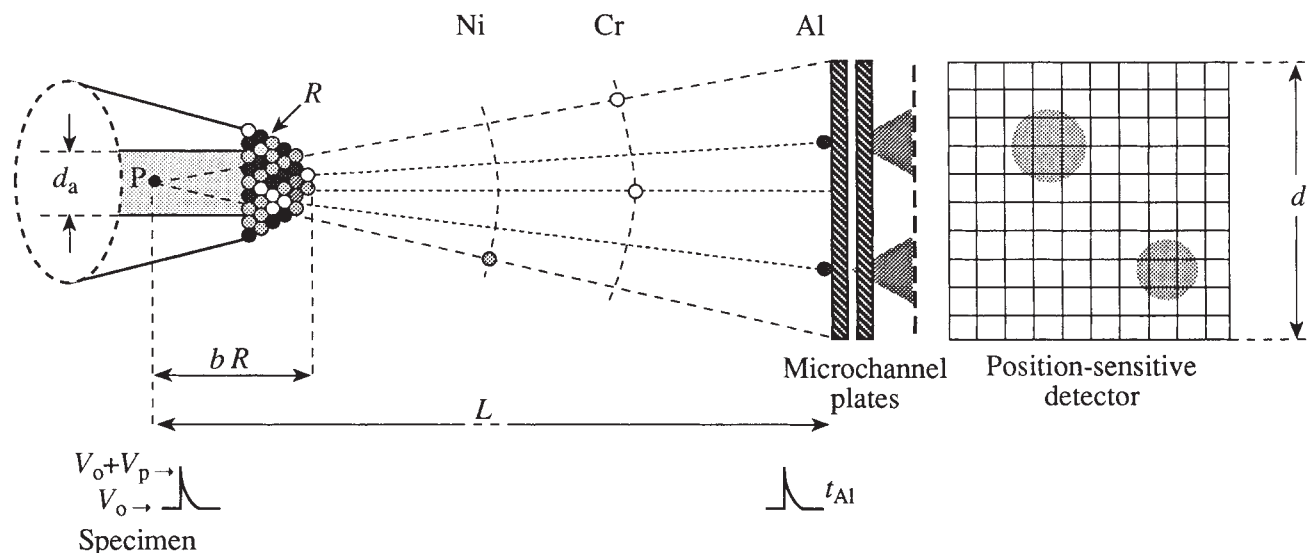


FIG. 1 Schematic view of the tomographic atom probe. The coordinates of atoms on the surface of a sample (a sharply pointed tip) are deduced from the position of ion impacts on the detector by means of a simple proportional operation. The magnification is $G = L/bR$. For a pure stereographic projection, the ion trajectories intercept at a single point (P) which is located at $2R$ from the tip surface ($b = 2$).

selected region of the tip are received on a position-sensitive detector placed at a given distance (L) from the specimen. The mass to charge ratio (m/n) is deduced from the time of flight and the total applied voltage ($V_o + V_b$) as follows

$$mv^2/2 = ne(V_o + V_p) \text{ with } v = L/t$$

The basic principle of the instrument lies in the use of a multi-impact spatial detector. For each evaporation pulse (~ 100 pulses per second), this system must measure the time of flight and the coordinates of ion impacts. The position measurement has to be carried out independently for every arrival time (t_{Al} , t_{Cr} , t_{Ni} . . .). The idea here is to obtain successive ‘material images’ related to each chemical species with its particular value of m/n . Once the coordinates of impacts are known, the positions of atoms at the tip surface can readily be calculated from a simple geometrical projection (Fig. 1). The magnification G is proportional to the ratio of flight path length to tip radius: $G \approx L/bR$, where b (~ 2) is a constant related to the projection features. For $R = 50$ nm and $L = 1$ m, $G \approx 10^7$. As a 0.1-nm precision at the tip surface is aimed for, impact positions must be determined with a 1-mm accuracy. The position-sensitive detector is composed of two chevron microchannel

Field-evaporated ions originating from the specimen surface are identified by time-of-flight mass spectrometry. Their positions are deduced from the coordinates of ion impacts on the detector. For each impact, an electron shower is produced by the microchannel plates, and the measurement of this charge received on a 10×10 anode array allows the position of the spot centre to be determined.

plates in front of a collector array of a hundred square nickel anodes (each 1 cm^2) deposited on a glass support. The detector size ($10\text{ cm} \times 10\text{ cm}$) leads, for $G \approx 10^7$, to an analysed area close to $10\text{ nm} \times 10\text{ nm}$. The geometry of this detector is similar to our earlier four-sector detection system⁸.

The calculation of coordinates relies on the measurement of the charge spread over the anode array⁹. Each ion impact on the microchannel plates generates a charge cloud which irradiates a few anodes. The spot diameter is typically 1–2 cm, depending on the number of electrons collected.

By measuring the electronic charges on each individual anode, we can calculate the positions of ion impacts. The principle is depicted in Fig. 2, which shows the cloud impinging on four anodes. When the charge on the right-hand anodes (Q_r) is equal to the charge on the left hand anodes (Q_l), the electron cloud is centred at $x = 0$ (local axes). As the impact shifts towards the right side, the reduced charge $q_x = Q_r/Q_t$ (with $Q_t = Q_r + Q_l$) will increase; its value is related to the position of the ion impact along the x -axis. The y -coordinate is determined similarly. As the anode size is 1 cm, the 1 mm precision which is required may be easily obtained. The precise procedure used to determine the transfer functions $x(q_x)$ and $y(q_y)$ and the

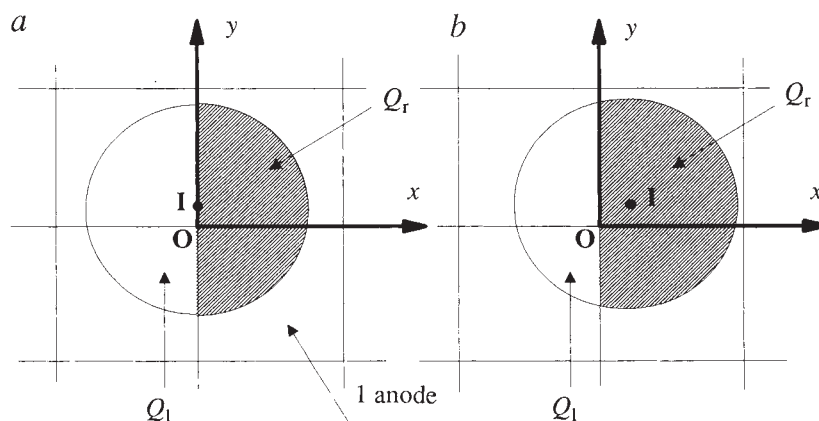


FIG. 2. The coordinates (x, y) of the ion impact (I) may be deduced from the elemental electronic charges collected on the multi-anode. The electron spot (dashed lines) produced by ion impact on the microchannel plate must irradiate several anodes. *a*, In this example, when the charge on the right side (Q_r) is equal to that on the left side (Q_l), the centre of the spot is at $x = 0$ with respect to the local axis shown. For each value of the reduced charge $q_x = Q_r / Q_t = (Q_t = Q_r + Q_l)$, one can calculate a single value of x . *b*, Here, $q_x > 1/2$ and $x > 0$.

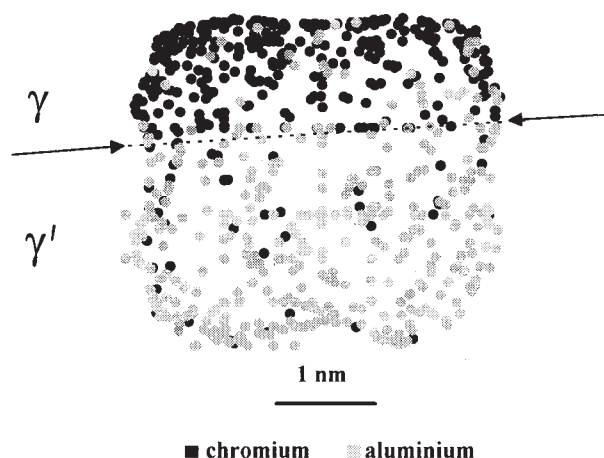


FIG. 3 Two-dimensional spatial distribution of chromium and aluminium atoms in a nickel-base superalloy (the nickel atoms are omitted). This map shows the boundary between an aluminium-rich γ' particle and the aluminium-depleted matrix. Only a portion of the precipitate is visible here. The transition (interface) between this particle and the surrounding chromium-rich matrix takes place over less than 0.5 nm.

technical details of the multidetector are described elsewhere⁷. To obtain the positions of each chemical species, these charge measurements are done separately for every arrival time. As ions with similar m/n ratios may be separated by a few tens of nanoseconds, fast electronics must be used. In the detection system, up to eight charge maps may be captured for each high-voltage pulse. Several ions with similar mass-to-charge ratios can be localized for each map. Under standard detection rate conditions (~ 1 ion per voltage pulse on average for the overall anode array), up to eight different times of flight may be detected. As a result, up to 800 measurements of charge (8×100 anodes) may be made on each evaporation pulse.

A two-dimensional map of aluminium and chromium in a nickel-base superalloy is shown in Fig. 3. This two-phase material contains aluminium-enriched γ' precipitates in a

chromium-rich γ solid solution. The calculated position of each individual atom is given with a precision of a few ångströms. Each anode corresponds to a spatial extent close to 1 nm. The map clearly shows the spatial distribution of elements and the presence of a γ/γ' interface. The phase boundary is very narrow (<0.5 nm), demonstrating that the instrument can reveal features of the material microstructure with subnanometre precision. The uncertainties are essentially controlled by the physical limits, namely the geometrical aberrations of ion trajectories.

The mass spectra of both γ and γ' phases are shown in Fig. 4. The mass resolution is fairly good (full width at half maximum ~ 200). This is substantially less than that of energy-compensated atom probes, but considerably better than the resolution of the position-sensitive atom probe PoSAP⁶. The measured composi-

FIG. 4 Mass spectra for γ' precipitates (a) and γ matrix (b). Ions were detected mainly in the doubly charged state (Ti^{2+} , Cr^{2+} , Ni^{2+} , Co^{2+}), except aluminium which is observed in both 2+ and 3+ charge states. The composition of the γ matrix as measured with the tomographic atom probe agrees well with the expected concentration levels as given by ref. 10. Precision is limited by the counting errors ($\sim 2\sigma = 2\sqrt{C(1-C)/N}$, where C is the concentration and N the number of collected ions; here $N = 10,000$). The elemental ratios determined by the atom probe are: Al: Ti: Cr: other (Ni, Co) = $4.1 \pm 0.4 : 0.14 \pm 0.08 : 24.3 \pm 0.9 : 71.5 \pm 0.9$ (2σ deviations). The expected composition¹⁰ is 3.1 : 0.6 : 25.5 : 70.8.

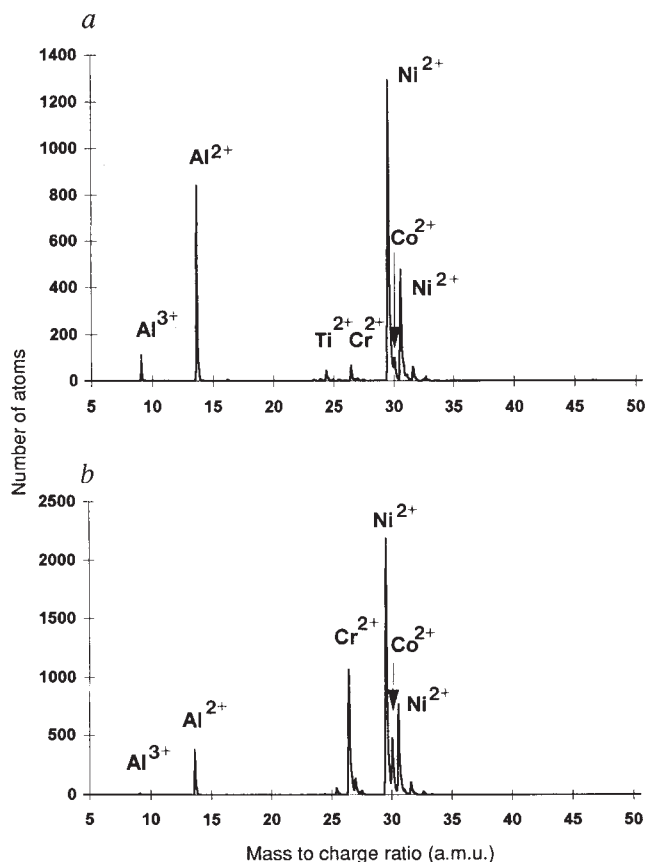
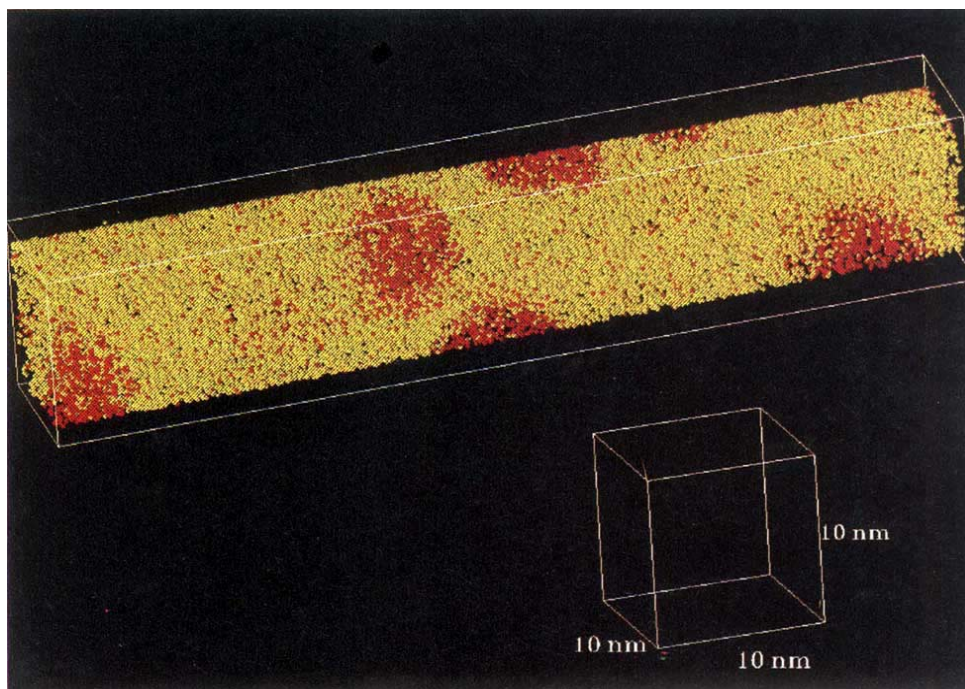


FIG. 5 Three-dimensional reconstruction of a small volume of a nickel-base superalloy containing small aluminium-enriched γ' particles (diameter ~ 7 nm) finely dispersed in a chromium-rich γ solid solution. Aluminium atoms are represented as red dots, chromium atoms as yellow dots. Other added elements (Ti, Co) as well as nickel atoms are not represented for clarity. The largest length of the analysed volume (depth) is parallel to the tip axis. The analysis area (10×10 nm²) is centred on the tip axis (see Fig. 1).



tion of the γ solid solution agrees well with the expected composition¹⁰, if one takes into account the precision of measurements and the unavoidable concentration fluctuations which are known to occur in this material. This shows that quantitative composition data may be obtained with our instrument.

Images such as that displayed in Fig. 3 correspond to a layer a few ångströms deep. By field evaporation of surface atoms, the material can be investigated in depth, layer by layer. For each evaporated layer, a two-dimensional element map can be calculated, and a three-dimensional reconstruction of the probed volume can be achieved by combining these. The probed depth can easily be deduced from the number of layers removed. A three-dimensional reconstruction of a small volume of a nickel-base superalloy is shown in Fig. 5. One hundred and fifty thousand ions were collected during the experiment. The image shows the presence of very fine aluminium-enriched spherical γ' particles embedded in a γ chromium-rich matrix. The average diameter of precipitates is 7 nm. Again, the γ/γ' interfaces are sharp. Once this type of three-dimensional image has been obtained, local compositions can also be extracted from a selected region.

Our tomographic atom-probe may be considered as being nearly equivalent to a hundred conventional atom-probes working in parallel. The instrument is able to provide three-dimensional information: topological data and very-fine-scale microstructural features are rendered on a near-atomic scale and can be related directly to local compositions. Although improvements remain to be made to the detector, our tomographic atom-probe is already a powerful microanalytical tool. □

Sea level and initiation of Predynastic culture in the Nile delta

Daniel Jean Stanley & Andrew G. Warne*

Mediterranean Basin Program, Paleobiology E-207 NMNH, Smithsonian Institution, Washington DC 20560, USA

PREDYNASTIC occupation of the Nile valley and delta, dating back to at least 5000 BC, has been attributed by most archaeologists to environmental factors, primarily regional climate change and fluctuating Nile flood stages¹⁻⁵. Here we propose instead that initiation of farming settlements in the Nile delta was closely related to eustatic sea level. We present geological analyses of late Quaternary subsurface sections throughout the delta which reveal that the deceleration in sea-level rise that occurred at about 6500–5500 BC was a prime factor in the accumulation of Nile silt, and the creation of the widespread and fertile delta plain. As rising sea level reduced the gradient of the river course, a system of meandering Nile distributaries evolved, with increased overbank deposition burying the former (early Holocene), partially vegetated sandy plain. The broadening, seasonally flooded, fecund plain, with its increasing plant cover, provided a setting that was conducive to evolving agricultural activity and was therefore instrumental in the development of Predynastic communities in the Nile delta.

Cultivation and domestication of livestock developed as early as 8,000 years ago in the eastern Sahara west of the Nile valley^{6,7}. Agriculture, however, was not widespread along the Nile until after the sixth millennium BC. During the Predynastic period (before the First Dynasty of the pharaohs, from ~ 5000 to 3100 BC), farming and herding replaced hunting and gathering. Predynastic sites include dwellings and graves with evidence of

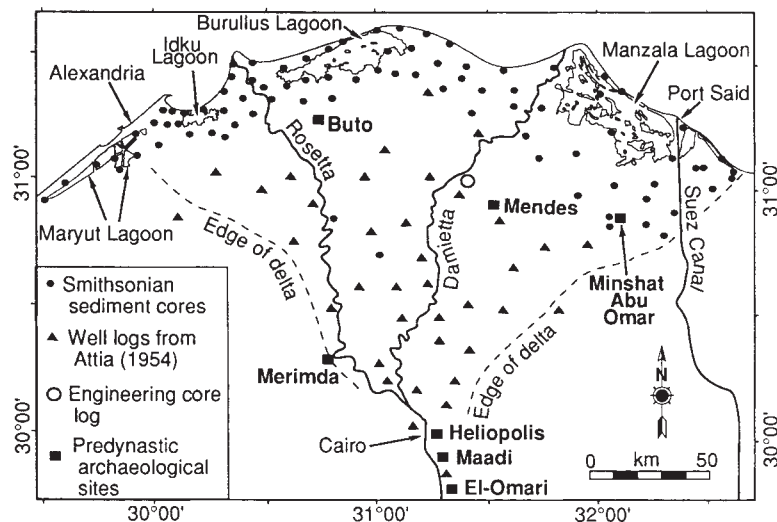
Received 1 March; accepted 16 April 1993.

1. Müller, E. W., Panitz, J. A. & McPane, S. B. *Rev. scient. Instrum.* **39**, 83–86 (1968).
2. Blavette, D. *Surf. Sci.* **266**, 299–309 (1992).
3. Miller, M. K. & Smith G. D. W. *Atom-Probe Microanalysis: Principle and Applications to Materials Problems* (Mat. Res. Soc., Pittsburg PA, 1989).
4. Panitz, J. A. *J. Vac. Sci. Technol.* **11**, 206–210 (1974).
5. Miller, M. K. *Surf. Sci.* **266**, 494–499 (1992).
6. Cerezo, A., Godfrey, T. J. & Smith, G. D. W. *Rev. scient. Instrum.* **59**, 862–866 (1988).
7. Deconihout, B. *et al. Appl. Surf. Sci.* (in the press).
8. Blavette, D., Bostel, A. & Sarrau, J. M. J. *Phys.* **47**, 473–478 (1986).
9. Wijnaendts Van Resandt, R. W., Harink, H. C. & Los, L. J. *Phys. E: Scient. Instr.* **9**, 503–509 (1976).
10. Blavette, D. & Bostel, A. *Acta met.* **32**, 811–816 (1984).

ACKNOWLEDGEMENTS. We thank the DRET, EDF, the CNRS, the MRT and the Région de Haute-Normandie for financial support.

* Present address: U.S. Army Corps of Engineers Waterways Experiment Station (CEWES-GG-YH), Vicksburg, Mississippi 39180-6199, USA.

FIG. 1 Map of the Nile delta, Egypt, showing sediment-core data base used for this study, and location of selected Predynastic sites.



domesticated wheat, barley, goats, sheep, cattle, pigs and artefacts such as ground stone and pottery³. It is generally held that widespread development of communities during this period led to evolution of Egyptian civilization⁴.

Merimda Beni Salama⁸, the oldest site in the delta north of Cairo (Fig. 1), lies on the western delta margin⁹. Other Predynastic settlements include Heliopolis, Maadi and El-Omari just south of the delta; sites in Middle and Upper Egypt include the Fayoum, Badari, Abydos, Nagada and Hierakonpolis³⁻⁵. Several Predynastic sites have been uncovered from beneath the mantle of Nile silts in the delta

proper. Settlements, for the most part younger than Merimda, include Minshat Abu Omar, Buto¹⁰ and Mendes¹¹; sites were positioned along water courses, generally on or near sand and silt hills termed *geziras*². These topographic highs served as refuge during annual floods, yet were accessible to channels and fertile floodplains necessary for grazing, cultivation and fishing. Some Predynastic sites were subsequently removed by erosion and/or have been destroyed by man³. It is likely that other settlements in the lower-lying portions of the delta plain, as yet undetected, have been buried by Nile alluvium in the course of floodplain aggradation⁹.

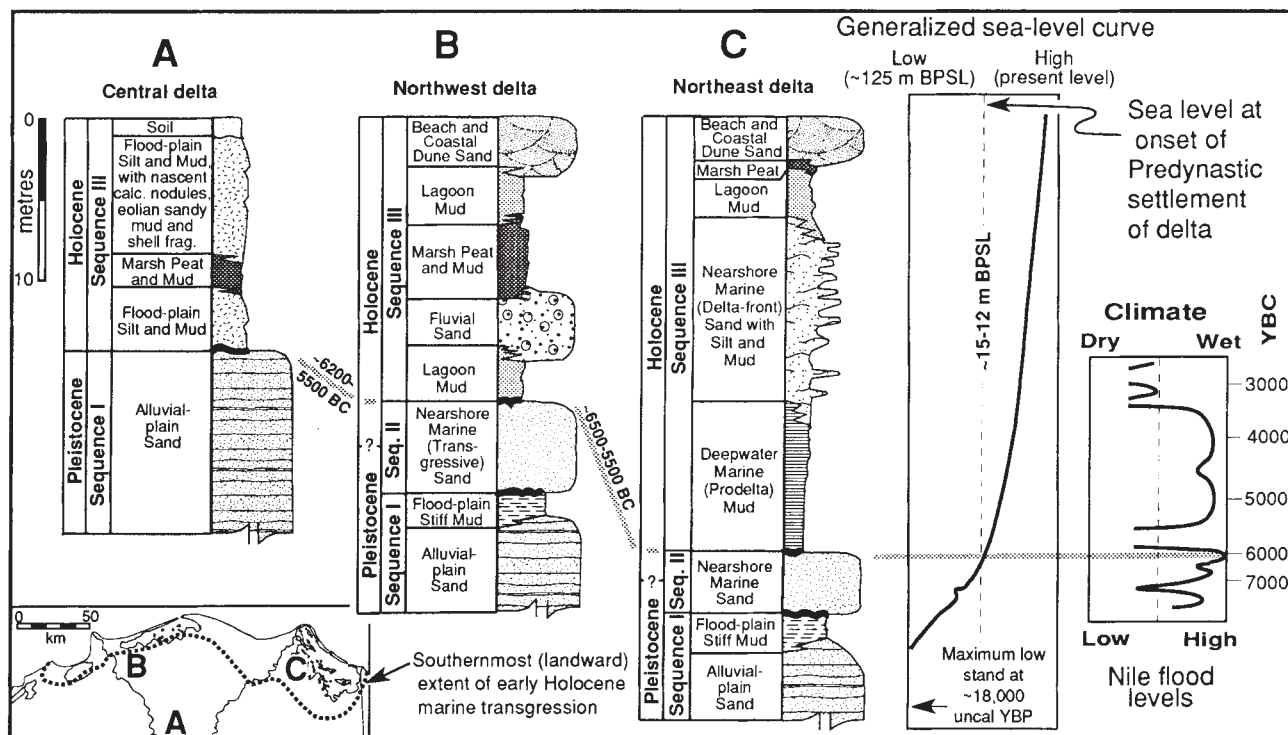


FIG. 2 Summary of late Quaternary sedimentary facies and associated environments of deposition in three regions of the Nile delta (locations shown in inset). These radiocarbon-dated facies (see ref. 12) are divided into three lithostratigraphic sequences (I, II, III). The main change in sedimentation (from sand to mud) at about 6500 to 5500 bc correlates with marked deceleration in sea-level rise (curve

after refs 21, 22) and with a relatively brief period of aridification (curve after ref. 4). Onset of mud-rich sequence III shortly precedes development of Predynastic culture in the delta. BPSL, below present sea level; YBC, years before Christ; uncal YBP, uncalibrated (radiocarbon) years before present.

Archaeological investigations document the appearance of Predynastic sites in the Nile delta and valley between ~5000 and 4400 BC^{4,5}. Early development of Predynastic sites along the Nile seems to correspond with a period of aridity in the Sahara and declining Nile flood levels⁴. A period of severe droughts during the middle Holocene in the eastern Sahara has been suggested as the primary cause of immigration of farming communities into the Nile region^{1,4,5}.

We concur with archaeologists that environmental factors were instrumental in the onset of Predynastic settlements in the Nile delta during early to mid-Holocene times²⁻⁵. This is borne out by extensive subsurface geological exploration, carried out as part of the Smithsonian's Nile Delta Project since 1985¹². Petrological, geochemical and faunal analyses were made on 87 long (up to 50 m), continuous, radiocarbon-dated drill cores along the entire northern delta (Fig. 1). More than 3,500 petrological samples were studied and >340 peats and organic-rich sediments were radiocarbon-dated¹³, enabling us to define late Quaternary sedimentary facies and their geometry in time and space (Fig. 2). In order that our dated geological samples correspond with those of Egyptologists, radiocarbon dates younger than ~8,000 years BP are calibrated to calendar years BC (by Beta Analytic Inc.; see ref. 14).

Petrologically detailed cores in the northern delta are correlated with radiocarbon-dated borings¹⁵, engineering cores and published well logs¹⁶ in the central and southern delta (Fig. 1). The litho- and chronostratigraphy of late Pleistocene and Holocene sections are summarized in logs for three regions (Fig. 2). Three main stratigraphic sequences (coded from I to III, from base to top of the logs) are apparent: I, late Pleistocene alluvial sands with localized stiff muds; II, late Pleistocene to early Holocene shallow marine (transgressive) sands; and III, Holocene deltaic sands, silts and muds (as thick as 50 m) of marine to terrestrial origin¹². These three sequences are separated from each other by unconformities (gaps in sedimentation). Subsurface sections, dated from >30,000 years BP to present, clearly record the abrupt and distinct, delta-wide change, from sand-dominant (sequences I and II) to mud-rich (sequence III) deposits between ~6500 and 5500 BC. More than 60 radiocarbon-dated samples¹³ from the base of sequence III reveal that this change in sedimentary regime occurs ~500 years prior to the first agricultural communities in the delta.

The marked depositional change could be related to regional fluctuations in Holocene climate and associated flood levels. Such oscillations are substantiated by studies in Upper Egypt and further to the south^{4,17-20}. One such curve for climate and flood stages (ref. 4 in Fig. 2) reveals an important arid/low-flood stage at about 6000–5500 BC. Fine-grained sediment began to accumulate in the Nile delta (sequence III, Fig. 2) just before the above-cited arid phase (~6500 to 5500 BC). However, depositional patterns during the past 8,000 years¹² clearly indicate that climatic and flood-level fluctuations (on the order of 200–1,000 years) have not been the only controlling factors on Holocene sedimentation in the Nile delta. If climate was the primary factor, we would expect sequence III to be cyclic, with marked alternating aggradational-degradational alluvial deposits in the cores. Detailed analyses of sequence III, however, reveals that after ~6000 BC there has been generally continuous aggradation of silt-rich sequences which are marine-influenced in the northern delta and fluvio-deltaic in the central and south delta (Fig. 2).

Palaeogeographic reconstructions of the delta demonstrate that aggradation and progradation of the Nile depocentre (as recorded by sequence III) was mostly caused by change in sea level rather than by regional climate factors alone (Fig. 2). Sea level dropped as much as 125 m below its present stand about 20,000–18,000 years ago²¹. At that time, the Nile shoreline was located near the present shelf break, as much as 50 km north of the present coast (Fig. 3a). This low sea-level stand resulted in a lowered river base level and incision of fluvial channels into the

alluvial plain, producing an increased Nile gradient of 1:1,450 between Cairo and the shelf edge. During this time, coarse sediment (largely sand) accumulated on the subaerially exposed plain and continental shelf, and fine sediment was bypassed to the shelf-edge delta, north of Alexandria¹⁵. The present delta region during this period was a sandy (sequence I, Fig. 2), rather barren and partially vegetated, steppe-like region (Fig. 4a). This plain of low relief (<15 m) included localized sand hills and dunes, and was traversed by a network of braided channels. These were interspersed with shallow evaporitic depressions, or *sebkhas*, subject to seasonal flooding and desiccation¹⁵.

Sea level rose rapidly (rates to 9 mm yr⁻¹), from ~125 to 16 m below its present stand, between about 18,000 and 8,000 years ago^{21,22}. This fast rise brought about an abrupt landward shift of the coastline, from near the shelf break to the northern sector of the present delta plain (Fig. 3b). This rise elevated the river base level so that channels became less incised and gradient decreased to 1:4,000 between Cairo and the coast. Fine sediment, however, continued to bypass the delta plain, and intense reworking by transgressing shallow marine waves and currents removed silt and clay from the northern delta region. Bypassing and reworking resulted in accumulation of moderately sorted transgressive sands (sequence II, Fig. 2) on the Egyptian shelf and northern delta plain (landward limit shown on inset in Fig. 2). During this period, the region was not yet conducive to development of farming settlements: although climate at times was more moist than today, the subaerially exposed, sandy plain south of the coast (Fig. 4b) remained partially vegetated and suffered periodic, destructive flooding.

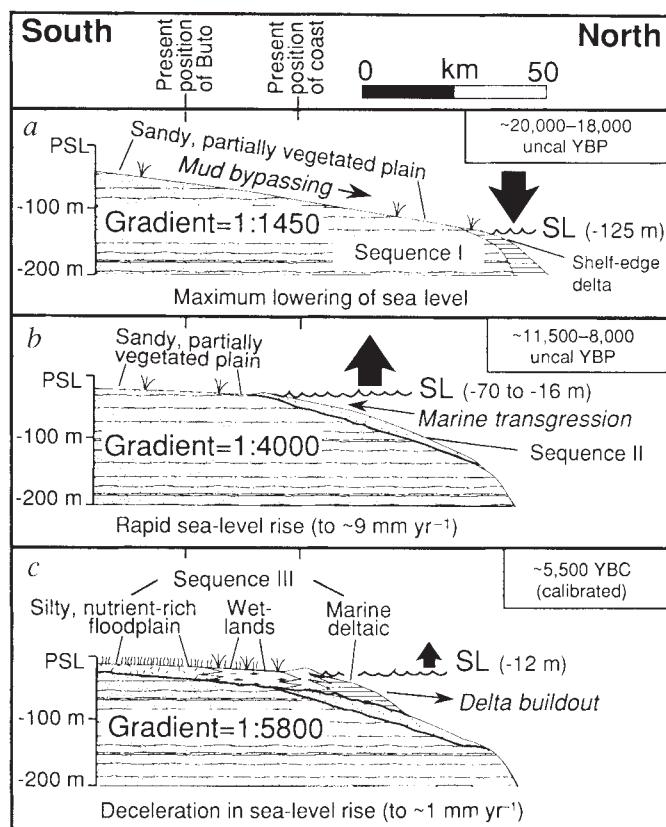


FIG. 3 Schematic north-south sections across the mid- and lower Nile delta plain and contiguous continental shelf during three phases of eustatic late Quaternary sea-level change: a, at maximum low stand, giving rise to deposition of sequence I; b, period of rapid rise in sea level, and deposition of sequence II; and c, during deceleration in sea-level rise, and onset of accretion of mud-rich sequence III. Further explanation in text. SL, sea level; PSL, present sea level.

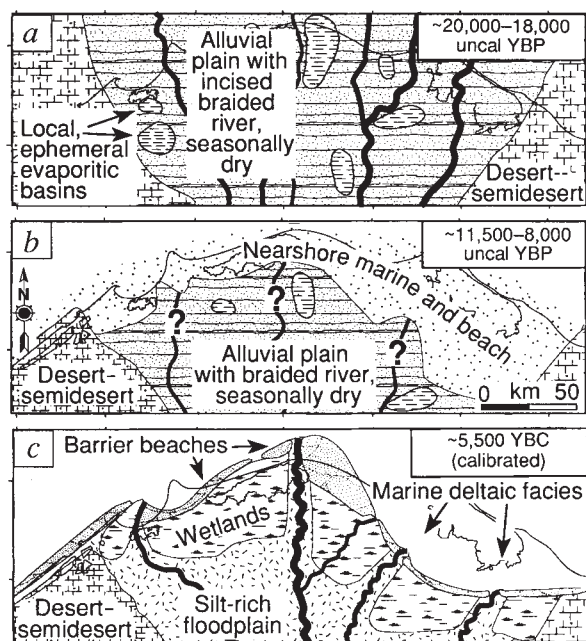


FIG. 4 Palaeogeographic maps of the northern and central Nile delta during the three phases of sea level depicted in Fig. 3. *a*, During lower sea-level stands, the northern delta and adjacent shelf was a subaerially exposed sandy, partially vegetated plain. *b*, By ~8,000 years ago, the sea had advanced onto the northern delta which became covered by reworked, shallow marine (transgressive) sand. *c*, At ~6500–5500 bc, sea-level rise decelerated, river gradients decreased, and a system of meandering Nile distributaries began to deposit large volumes of nutrient-rich overbank silts. The broadening, fertile floodplain was conducive to Predynastic agriculture.

The rate of sea-level rise decelerated markedly from about 7000 to 5000 BC, to $\sim 1 \text{ mm yr}^{-1}$ (refs 22, 23). As sea level slowly rose to about 12 m below present stand at ~6000 BC, base level was elevated and gradient decreased to 1:5,800 (Fig. 3c). During this period, the rate of sediment accumulation matched or exceeded sea-level rise (see also ref. 24), shoreline position became relatively stable (Fig. 4c), and formation of the modern Nile delta began (sequence III, Fig. 2). Lagoons and marshes developed in the northern third of the delta, landward of coastal barrier beaches. South of this wetland region, rising base level and decreasing gradient gave rise to a meandering distributary channel system which, during annual Nile floods, deposited large volumes of nutrient-rich overbank silts. These accreting fine-grained deposits formed a broadening, fertile floodplain which retained enough moisture for longer periods during the year to allow plants to flourish. Such an increase in floral activity in the delta is indicated by palaeobotanical studies. Basal Holocene core sections at ~6000 BC contain abundant grasses and other plants^{25,26} indicating that Nile delta silts could sustain grazing and cultivation, even before artificial irrigation². Moreover, cores record prolific plant growth, even during times

of aridity, at about 6000 BC and during subsequent periods (Fig. 2). As accretion progressed and larger areas of the plain were mantled with silt, people moved into inner parts of the delta (for example to Buto and Minshat Abu Omar).

The Late Pleistocene to Holocene lithological changes that we observe in the Nile are comparable to those in the subsurface of many of the world's major deltas. Examples include the Rhône²⁷ and Ebro²⁸ deltas in the Mediterranean, and the Mississippi²⁹ and Yangtze³⁰ deltas in other oceans. All these depocentres show the distinct break from sandy, oxidized alluvial plain in the late Pleistocene to silty, nutrient-rich floodplain in the early- to mid-Holocene¹⁵. This major lithological transition, occurring under markedly different climate regimes, is best explained by the worldwide deceleration in sea-level rise at ~6000 BC. We conclude that it was primarily the onset of floodplain silt accretion, related mainly to sea-level deceleration rather than to regional climate fluctuations alone, that fostered the transition from hunting and gathering to agriculture in the delta. Further study is needed to determine the possible influence of sea-level rise and consequent elevation in river base level on cultural development further to the south along the Nile valley. □

Received 25 February; accepted 20 April 1993.

- Chiold, V. G. *New Light on the Most Ancient East*, 4th Edn (Grove, New York, 1957).
- Butzer, K. W. *Early Hydraulic Civilization in Egypt* (Univ. of Chicago Press, Chicago, 1976).
- Hoffman, M. A. *Egypt Before the Pharaohs* (Dorset, New York, 1979).
- Hassan, F. A. J. *World Hist.* **2**, 135–185 (1988).
- Wenke, R. J. J. *World Prehist.* **5**, 279–329 (1991).
- Wendorf, F., Schild, R. & Close, A. E. (eds) *Cattle-Keeper of the Eastern Sahara* (Southern Methodist Univ., Dallas, 1984).
- Wendorf, F. *et al.* *Nature* **359**, 721–724 (1992).
- Eiwanger, J. *Mitt. Dtsch. Archäol. Inst. Abt. Kairo* **38**, 67–82 (1982).
- Hawass, Z., Hassan, F. A. & Gautier, A. J. *Egypt. Archaeol.* **74**, 31–38 (1988).
- Andres, W. & Wunderlich, J. in *The Nile Delta in Transition: 4th–3rd Millennium BC* (ed. van den Brink, E. C. M.) 157–166 (Pinkhas, Tel Aviv, 1992).
- Brewer, D. J. & Wenke, R. J. in *The Nile Delta in Transition: 4th–3rd Millennium BC* (ed. van den Brink, E. C. M.) 191–198 (Pinkhas, Tel Aviv, 1992).
- Stanley, D. J. & Warne, A. G. *Science* **260**, 628–634 (1993).
- MEDIBA (Mediterranean Basin Program) *Nile Delta Project Data-base Listings* (U.S. National Museum of Natural History, Washington DC, 1993).
- Stuiver, M. & Kra, R. *Radiocarbon* **28**, 805–1030 (1986).
- Chen, Z. & Stanley, D. J. *J. Coast. Res.* **9**, 539–576 (1993).
- Attia, M. *Deposits in the Nile Valley and the Delta* (Geological Survey of Egypt, Cairo, 1954).
- Adamson, D. A., Gasse, F., Street, F. A. & Williams, M. A. J. *Nature* **288**, 50–55 (1980).
- Said, R. *The Geological Evolution of the River Nile* (Springer, New York, 1981).
- Degens, E. T. & Spitz, A. *Mitt. Geol. Paläontol. Inst. Univ. Hamburg SCOPE/UNEP* **55**, 1–20 (1983).
- Paulissen, E. & Vermeersch, P. M. *Bull. Soc. Géol. Fr.* **5**, 73–83 (1989).
- Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
- Lighty, R., Macintyre, I. & Stuckenrath, R. *Coral Reefs* **1**, 125–130 (1982).
- Emery, K. O. & Aubrey, D. G. *Sea Levels, Land Levels, and Tide Gauges* (Springer, New York, 1991).
- Scruton, P. C. in *Recent Sediments, Northwest Gulf of Mexico* (eds Shepard, F. P., Phleger, F. D. & van Andel, T. J. H.) 82–102 (Am. Assoc. Petrol. Geol., Tulsa, 1960).
- Saad, S. I. & Sami, S. *Pollen Spores* **9**, 467–503 (1967).
- Leroy, S. A. G. *Vegetat. Hist. Archaeobot.* **1**, 43–52 (1992).
- Oomkens, E. in *Deltaic Sedimentation, Modern and Ancient* (ed. Morgan, J. P.) *Soc. Econ. Paleontol. Mineral. spec. Publ.* **15**, 198–212 (1970).
- Maldonado, A. in *Deltas: Models for Exploration* (ed. Broussard, M. L.) 311–338 (Houston Geological Society, Houston, 1975).
- Coleman, J. M. *Bull. Geol. Soc. Am.* **100**, 999–1015 (1988).
- Yang, H. & Xie, Z. *Oceanol. Limnol. Sin.* **15**, 1–13 (1984).

ACKNOWLEDGEMENTS. We thank members of the Nile Delta Team for assistance in compiling and interpreting data from Nile delta cores. F. A. Hassan, I. G. Macintyre, T. H. van Andel and R. J. Wenke provided suggestions which improved the manuscript. Funding was provided by the Smithsonian (Scholarly Studies Program, Assistant Secretary for Science, NMNH Director's Office, Sprague Fund) and Elf Aquitaine-Washington.

Vegetation effects on the isotope composition of oxygen in atmospheric CO₂

Graham D. Farquhar*, Jon Lloyd*, John A. Taylor†, Lawrence B. Flanagan‡, James P. Syvertsen§, Kerry T. Hubick*, S. Chin Wong* & James R. Ehleringer||

* Plant Environmental Biology Group, Research School of Biological Sciences, Institute of Advanced Studies, Australian National University, GPO Box 475, Canberra, ACT 2601, Australia

† Centre for Resource and Environmental Studies, Institute of Advanced Studies, Australian National University, Canberra, ACT 2601, Australia

‡ Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa K1S 5B6, Ontario, Canada

§ Citrus Research and Education Center, University of Florida, Lake Alfred, Florida 33850, USA

|| Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

THE ¹⁸O/¹⁶O ratio in atmospheric CO₂ is a signal dominated by CO₂ exchange with the terrestrial biosphere and it has considerable potential to resolve the current importance of the oceans and individual terrestrial biomes as net sinks for anthropogenic CO₂. Fractionation of the oxygen isotopes of CO₂ occurs in plants owing to differential diffusion of C¹⁸O¹⁶O and C¹⁶O₂ and to isotope effects in oxygen exchange with chloroplast water. Here we investigate the consequences of these effects for the global distribution of oxygen isotopes in CO₂. We predict that ¹⁸O isotopic exchange fluxes, especially between the atmosphere and terrestrial biosphere, are large, with considerable spatial variation. Near 70° N, where precipitation (and soil water) is most depleted in ¹⁸O, photosynthesis and respiration both deplete the atmospheric CO₂ of ¹⁸O. This provides an explanation for the depletion of ¹⁸O in atmospheric CO₂ at high northern latitudes¹.

During terrestrial photosynthesis, heavier C¹⁸O¹⁶O molecules diffuse into leaves more slowly than do the lighter C¹⁶O₂. Further, within the chloroplasts, carbonic anhydrase catalyses the exchange of oxygen atoms between CO₂ and the water there. The ¹⁸O/¹⁶O ratio of chloroplast water is enriched compared with soil water because H₂¹⁸O evaporates more slowly from leaves than does H₂¹⁶O. Some of the CO₂ entering the chloroplast and exchanging oxygen atoms with chloroplast water is assimilated via photosynthesis, but some diffuses back into the atmosphere, the amount depending on the concentration of CO₂ in the chloroplast and on conductances to diffusion within and from the leaf. CO₂ molecules diffusing out of the leaf after exchanging oxygens are usually, but not always, enriched in ¹⁸O compared to those in the ambient air. CO₂ released in respiration and decomposition is less enriched because it largely reflects CO₂ in equilibrium with unenriched soil water.

The ¹⁸O composition of water at the sites of evaporation within leaves, δ_E, is enriched with respect to source (soil) water. Following the derivation for a free water surface², an equation for δ_E can be presented³

$$\delta_E = \delta_S + \varepsilon_k + \varepsilon^* + (\delta_v - \delta_S - \varepsilon_k) \frac{e_a}{e_i} \quad (1)$$

where δ_S is the isotope composition of source water, δ_v is the isotopic composition of water vapour in the surrounding air, *e_a* and *e_i* are the vapour pressures in the atmosphere and intercellular spaces, ε* is the proportional depression of equilibrium vapour pressure by the heavier molecule (9.2% at 25 °C; ref. 4) and ε_k is the kinetic fractionation factor, 28.5% for diffusion through stomata⁵ and 18.9% in the boundary layer³. Liquid H₂¹⁸O is concentrated at the evaporating sites and so diffuses away from them in the liquid phase, relative to H₂¹⁶O. This is

Symbols used in text

$\bar{a}$	Weighted mean of discriminations occurring during the diffusion of CO ₂ from ambient air to the sites of carboxylation
a_w	Fractionation against C ¹⁸ O ¹⁶ O (compared to C ¹⁶ O ₂) during dissolution and diffusion in water
<i>A</i>	Net rate of CO ₂ assimilation (=GPP at regional or global level)
<i>c_a</i>	Mole fraction of CO ₂ in air
<i>C_a</i>	Atmospheric partial pressure of CO ₂ (<i>p</i> CO ₂)
<i>C_c</i>	<i>p</i> CO ₂ in chloroplast
<i>C_{st}</i>	<i>p</i> CO ₂ in substomatal cavities
<i>D</i>	Leaf-to-air vapour pressure difference
δ _a	Oxygen (O) isotope composition of atmospheric CO ₂ relative to PDB-CO ₂
δ _{an}	O isotope composition of CO ₂ released anthropogenically (PDB-CO ₂ scale)
δ _c	O isotope composition of CO ₂ in chloroplast (PDB-CO ₂ scale)
δ _E	O isotope composition of water at sites of evaporation in leaf (SMOW scale)
δ _L	O isotope composition of bulk leaf water (SMOW scale)
δ _o	O isotope composition of water in the ocean (SMOW scale)
δ _p	O isotope composition of precipitation (SMOW scale)
δ _r	O isotope composition of respired CO ₂ (PDB-CO ₂ scale)
δ _s	O isotope composition of water in soil (SMOW scale)
δ _v	O isotope composition of water vapour in air (SMOW scale)
Δ _A	Discrimination against ¹⁸ O (compared to ¹⁶ O) during net CO ₂ assimilation
<i>e_a</i>	Ambient water vapour pressure
<i>e_i</i>	Water vapour pressure in intercellular spaces of leaves
<i>E_v</i>	Elevation
ε _k	Kinetic fractionation against H ₂ ¹⁸ O compared to H ₂ ¹⁶ O during diffusion in air
ε*	Proportional depression of vapour pressure of H ₂ ¹⁸ O compared to H ₂ ¹⁶ O
<i>F_{an}</i>	Molar flux of CO ₂ to atmosphere from anthropogenic sources
<i>F_{ao}</i>	Gross (one way) molar flux of CO ₂ from atmosphere to ocean
<i>F_{oa}</i>	Gross molar flux of CO ₂ from ocean to atmosphere
φ _c	Proportion of <i>A</i> lost in plant respiration
<i>g</i>	Conductance to diffusion of CO ₂ from atmosphere to chloroplasts
GPP	Gross primary productivity (molar carbon flux)
Γ	CO ₂ compensation point
λ	Lagrange multiplier representing the marginal water cost of plant carbon gain
<i>M</i>	Number of moles of air in the atmosphere
NPP	Net primary productivity (molar carbon flux)
<i>P_a</i>	Annual precipitation
PDB-CO ₂	CO ₂ formed by addition of phosphoric acid to Pee Dee Belemnite calcite
ℛ	Molar flux of CO ₂ released by plant and soil respiration
<i>R_a</i>	¹⁸ O/ ¹⁶ O ratio of CO ₂ in ambient air
<i>R_A</i>	¹⁸ O/ ¹⁶ O ratio of the net flux of CO ₂ into the leaf
SMOW	Standard mean ocean water
<i>T</i>	Annual mean temperature
<i>T_l</i>	Leaf temperature
<i>T_m</i>	Monthly mean air temperature

opposed by convection of source water into the leaf, a Péclet effect. Therefore bulk leaf water, δ_L, is less enriched in ¹⁸O than water at the evaporating sites. Most solar energy absorption occurs in chloroplasts and evaporation is from the adjacent air-water interfaces⁶. The ¹⁸O composition of water in the chloroplasts should therefore be closer to δ_E than to δ_L.

The enrichment in ¹⁸O in chloroplast water will be passed to CO₂, with interconversion being catalysed by the enzyme carbonic anhydrase with an equilibrium fractionation factor of 41.2‰ at 25 °C (ref. 7). The CO₂ can then be assimilated (at net rate *A*) or diffuse out of the leaf. The resulting isotope composition of CO₂ in the chloroplast, δ_c, affects the isotope discrimination against ¹⁸O during net CO₂ assimilation. We define this discrimination, Δ_A, as *R_a*/*R_A* - 1, where *R_a* is the ¹⁸O/¹⁶O ratio of ambient CO₂, and *R_A* is that ratio for the net flux of CO₂ into the leaf. This definition, based on the net rate of assimilation, *A* (that is, consumption of CO₂), is general enough to include possible effects of the carboxylating enzyme, but in what follows

we show that such effects are small and that the discrimination is effectively dominated by fractionation during diffusion and by the CO_2 -water exchange reaction.

Following analogous derivations⁸ for $^{13}\text{CO}_2$, we obtain an equation for ^{18}O discrimination in the case of complete isotopic equilibration between CO_2 and water before the fixation of CO_2

$$\Delta_A = \bar{a} + \frac{C_c}{C_a - C_c} (\delta_c - \delta_a) \quad (2)$$

where δ_a is the isotope composition of the ambient CO_2 , C_c and C_a are the partial pressures of CO_2 ($p\text{CO}_2$) in the chloroplast and ambient air, respectively, and $\bar{a}$ is the weighted mean of discriminations occurring during the diffusion from ambient air to the sites of carboxylation within the chloroplast. The value of $\bar{a}$ will be dominated by fractionation during diffusion from the leaf surface to the mesophyll cell wall (8.8‰, based on the reduced mass of CO_2 and air) but also includes smaller fractionations in the laminar boundary layer and during diffusion through solution. We estimate $\bar{a}$ to be ~7.4‰.

Figure 1 shows the relationship between Δ_A and C_c/C_a for leaves of C_3 tree species, for which on-line ^{13}C discrimination has been measured and analysed⁹. Also shown are the relationships between Δ_A and C_c/C_a for chloroplast water having the same composition as source water ($\delta_s = -8.0\text{‰}$ SMOW) and for water at the evaporating surface within the leaf. In both cases, equilibrium between CO_2 and H_2O is assumed. Values of Δ_A observed are close to CO_2 in full equilibrium with water at the sites of evaporation. Data shown in Fig. 1 are from woody species but similar results were obtained for wheat and soybean in Canberra and *Phaseolus vulgaris* in Utah. CO_2 produced from respiration and photorespiration can be less enriched than is expected for CO_2 in full equilibrium with δ_E (ref. 10). Nevertheless, isotope exchange during CO_2 assimilation is well predicted using equation (2). Δ_A can then simply be regarded as the net result of a flux, gC_a , into the leaf with an isotopic source δ_a (g

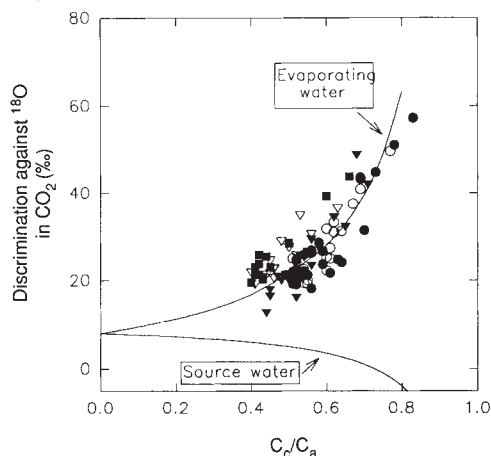


FIG. 1 Relationship between discrimination against $\text{C}^{18}\text{O}^{16}\text{O}$ (Δ_A) and the ratio of chloroplastic to ambient partial pressures of CO_2 (C_c/C_a) for C_3 tree species grown under conditions of either full sunlight or under shade cloth. Full experimental details related to ^{13}C discrimination have been provided⁹. The estimates of C_c used here are those obtained from ^{13}C discrimination, but with the internal conductance for CO_2 diffusion within the leaf multiplied by 0.7 to account for C_c estimates from $^{13}\text{CO}_2$ discrimination being weighted by assimilation rates, and C_c values appropriate for $\text{C}^{18}\text{O}^{16}\text{O}$ discrimination being weighted by diffusion conductances (see appendices 1 and 2 of ref. 9). ■, Sun peach (*Prunus persica*); ▼, shade peach; □, sun grapefruit (*Citrus paradisi*); ▽, shade grapefruit; ○, sun lemon. Also shown are the values expected if CO_2 were in full equilibrium with source water ($\delta^{18}\text{O}_{\text{SMOW}} = -8.0\text{‰}$) or in full equilibrium with water at the evaporating surface ($\delta^{18}\text{O}_{\text{SMOW}} = +7\text{‰}$ at 25°C and $D = 1.1\text{ kPa}$). $\delta^{18}\text{O}_{\text{PDB}}$ of CO_2 entering the leaf chamber was typically -7‰ . (Canberra).

being the conductance to diffusion), and a flux, gC_c , out of the leaf with the isotope source, δ_c , being equivalent to that for full equilibrium between CO_2 and evaporating surface water. Each flux is subject to discrimination, $\bar{a}$, during diffusion.

To calculate Δ_A on a global basis, we first estimated soil water isotope composition which reflects the precipitation (δ_p) with seasonal variability in δ_p almost completely dampened^{11,12}. It was calculated by regressing δ_p for sites in the IAEA network¹³ against annual mean temperature, T ($^\circ\text{C}$), annual precipitation, P_a (mm) and elevation, E_v (m). The best fit obtained was

$$\delta_p (\text{SMOW}) = (0.579T - 0.0114T^2 - 1.35P_a + 4.47P_a^2 - 0.147\sqrt{E_v - 9.80}) \times 10^{-3}$$

With δ_p simulated at 2.5° resolution using temperature¹⁴, precipitation¹⁵ and elevation¹⁶ fields, the result is shown in Fig. 2. To calculate δ_E , ϵ_k was taken as 26‰ and $(\delta_v - \delta_s)$ was calculated from simulations of the global H_2^{18}O atmospheric cycle¹⁷ for continental regions, taking monthly values of δ_v and annual precipitation-weighted values of δ_s . Leaf temperature (T_l) was calculated as $T_l = 1.05(T_m + 2.5^\circ)$. The 2.5° addition accounts for a daytime increase in air temperature over monthly mean air temperature (T_m) and the 5% allows for net canopy to air heat fluxes. Monthly relative humidities were calculated using dewpoint temperatures and adding 1°C to account for higher daytime vapour pressures.

We then used δ_a for 1985¹, constant at $+0.77\text{‰}$ between 90°S and 32°S , then decreasing linearly with latitude to -1.37‰ at 90°N . δ_c was taken as the value for CO_2 in full equilibrium with δ_E , and Δ_A was then obtained from equation (2) with $\bar{a} = 7.4\text{‰}$, allowing C_c/C_a to vary with leaf-to-air vapour pressure difference, D . We first calculated C_{st} , the $p\text{CO}_2$ in the substomatal cavities, using a simple stomatal model¹⁸ as $C_{st}/C_a = 1 - \sqrt{1.6D(C_a - \Gamma)/(\lambda C_a^2)}$, where Γ is the CO_2 compensation point and λ is a Lagrange multiplier¹⁹ dependent upon vegetation type and photosynthetic mode as defined by a 1° resolution data set of land cover²⁰. $(C_{st} - C_c)/C_a$ was taken as 0.1 (ref. 9). On a regional scale there are variations in temperature, δ_s (Fig. 2), relative humidity and C_c/C_a that cause Δ_A to vary (Fig. 3) from -20‰ in arctic tundra to $+32\text{‰}$ in the dry steppes of Kazakhstan and Ukraine. In the arid areas of Africa and Australia, Δ_A is low owing to the dominance of C_4 species with low C_c/C_a .

To assess the impact of discrimination against $\text{C}^{18}\text{O}^{16}\text{O}$ on a global scale, we then calculated a globally averaged Δ_A , weighted by CO_2 assimilation (gross primary productivity: GPP). Net primary productivity (NPP) was first related to temperature and precipitation^{21,22} but with NPP = 0 in months where mean air temperature was less than -5°C . A (=GPP) and NPP are related; $A = \text{NPP}/(1 - \phi_c)$, where ϕ_c is the proportion of A lost in plant respiration. We took into account the variation in ϕ_c , which ranges from 0.3 to 0.5 for herbaceous species^{23,24}, is 0.3–0.6 for temperate zone forests²⁵ and 0.65–0.80 for tropical forests²⁶. ϕ_c has been estimated and GPP values converted to a carbon basis²⁷ using land cover data²⁰. We estimate GPP to be $12.5\text{ Pmol C yr}^{-1}$ (where $1\text{ Pmol C} = 12 \times 10^9\text{ ton C}$), with 16% of the total by C_4 plants. Weighting according to GPP gives $\delta_s = -7.9\text{‰}$ (SMOW), $\delta_c = -18.2\text{‰}$ (SMOW), relative humidity at the leaf surface = 0.65, $\delta_l = +4.4\text{‰}$ (SMOW), $C_c/C_a = 0.57$, and $\Delta_A = 13.7\text{‰}$.

Uncertainty in C_c/C_a has the greatest potential impact on model results: a change of 1 p.p.m. in C_c will cause a change of almost 0.1‰ in the calculated Δ_A . The ratio C_c/C_a is, however, a reasonably conservative parameter, depending mostly on plant photosynthetic mode and D ^{6,18,19}. These effects are accounted for in the current model. Upon development of soil water deficits, C_c/C_a can decline^{6,19}. Nevertheless, under such circumstances A is low and hence contributes little to the GPP-weighted value calculated here. Furthermore, in productive regions that regularly experience soil water deficits, the vegetation is often domi-

nated by C_4 grasses. As can be seen from Fig. 1, Δ_A is less sensitive to C_c/C_a at lower values of that ratio, such as are typically found in C_4 plants. Plant $\delta^{13}C$ values can be used to estimate C_c of C_3 species^{8,9}. Further data from around the globe will help constrain this uncertainty.

The other significant source of uncertainty is $(\delta_v - \delta_s)$, with a 1% error in our estimate causing a 1% change in Δ_A for globally averaged conditions. The GPP-weighted value we calculate here (-10.3%) is close to that usually observed²⁸⁻³¹. Uncertainties in the estimation of leaf temperatures introduce only minor inaccuracies into the calculations, and estimates of Δ_A are relatively insensitive to the calculated relative humidity. This is because the stomatal model interacts with D . The reduction in C_c/C_a accompanying a decrease in atmospheric humidity counters the increase in δ_E .

With $GPP = 12.5 \text{ Pmol yr}^{-1}$ and an atmospheric CO_2 pool of 62.5 Pmol , we estimate a global exchange time of 5 years for the carbon atom, and with $C_c/C_a = 0.57$, only 2.2 years for the oxygen atoms in CO_2 . This is much faster than the exchange of oxygen atoms with the ocean (8.3 years). The ^{18}O content of atmospheric CO_2 is also influenced by release of CO_2 from soils and plants via respiration and release of CO_2 from fossil fuel and biomass burning^{1,32}. Therefore, as already done for $^{13}CO_2$ (ref. 33), we can write

$$Mc_a \frac{d\delta_a}{dt} = F_{oa}(\delta_o - a_w - \delta_a) + F_{ao}a_w + M(\delta_r - \delta_a) + A\Delta_A + F_{an}(\delta_{an} - \delta_o) \quad (3)$$

where M is the number of moles of air in the atmosphere, c_a

FIG. 2 Precipitation-weighted isotopic composition (SMOW) of precipitation.

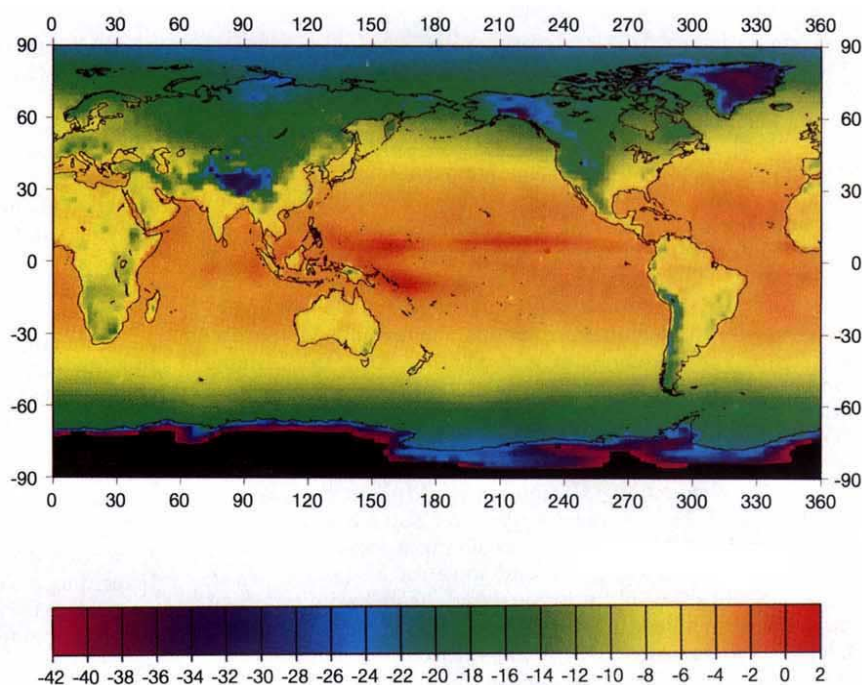
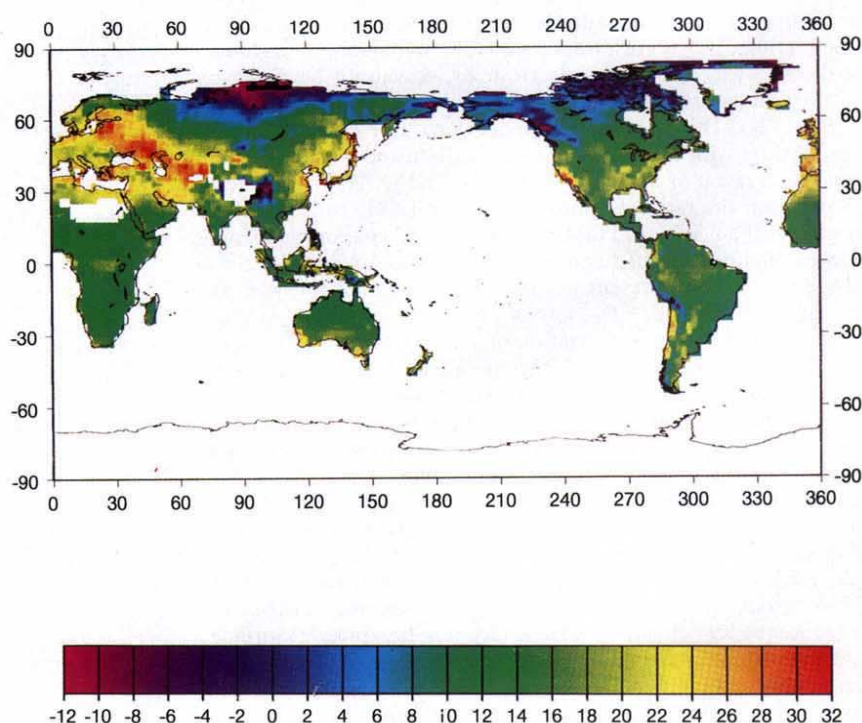


FIG. 3 Estimated mean annual values for discrimination against ^{18}O in CO_2 (Δ_A). Empty grid squares are modelled to have no terrestrial productivity.



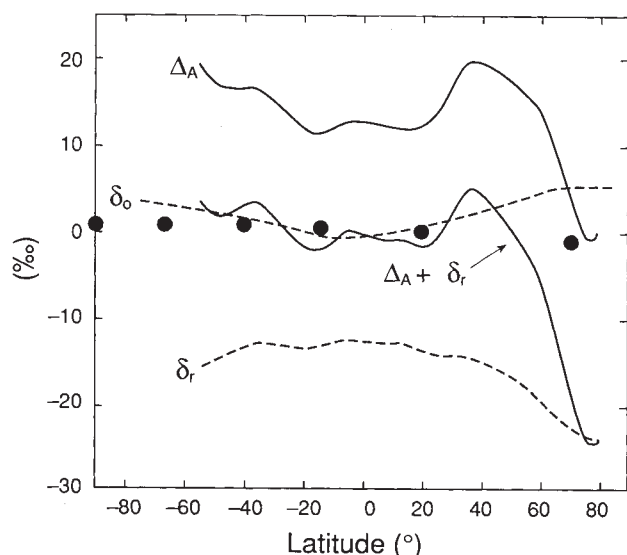


FIG. 4 10° latitudinal averages, weighted by GPP, of the predicted discrimination against ^{18}O in CO_2 (Δ_A), $\delta^{18}\text{O}$ (PDB- CO_2) of: respired CO_2 (δ_r), the sum $\Delta_A + \delta_r$, CO_2 in equilibrium with the ocean (δ_o) and mixed atmospheric CO_2 (δ_a data from ref. 1).

is the mole fraction of CO_2 in the atmosphere, F_{oa} and F_{ao} are the annual gross (one way) molar fluxes of CO_2 between oceans and atmosphere and atmosphere and oceans, $\mathcal{R}$ is the annual gross flux for CO_2 released from plant and soil respiration, A is the flux of CO_2 associated with photosynthetic CO_2 assimilation (gross primary productivity, GPP) and F_{an} is the influx of CO_2 from anthropogenic sources. δ_o is the isotopic composition of CO_2 in equilibrium with ocean water, δ_r is the isotopic composition of CO_2 respired by plants and soils as it enters the atmosphere, δ_{an} is the isotopic composition of anthropogenic CO_2 , and α_w is the combined solubility/aqueous diffusion fractionation for $\text{C}^{18}\text{O}^{16}\text{O}$.

Solution of equation (3) requires $d\delta_a/dt$, and the effective sink terms, $(F_{ao} - F_{oa})$ and $(A - \mathcal{R})$. We solve a simpler case (effectively pre-anthropogenic), assuming a steady state with $F_{oa} = F_{ao}$, $A = \mathcal{R}$, and $F_{an} = 0$, and solve for δ_r , taking³⁴ $F_{oa} = F_{ao} = 7.5 \text{ Pmol yr}^{-1}$. δ_o was analysed from salinity³⁵ and sea-surface temperatures³⁵ using surface-water salinity/ $\delta^{18}(\text{H}_2\text{O})$ relationships², calculating the isotopic composition of CO_2 (PDB- CO_2 scale) in equilibrium with ocean water (SMOW scale)^{36,37} at 2.5° resolution. This gives $\delta_o = +1.25\%$. Weighted according to cosine of latitude, $\delta_a = +0.18\%$. From inspection of equation (3), with respiration and GPP in balance on a global scale, because δ_a differs little from δ_o , δ_a should be close to $\Delta_A + \delta_r$. Indeed, from equation (3) we calculate that $\delta_r = -14.0\%$ (PDB- CO_2). This is 1.2% enriched compared with CO_2 in full equilibrium with soil water^{12,38}, -6.4% , followed by 8.8% depletion¹² as a result of slower diffusion of $\text{C}^{18}\text{O}^{16}\text{O}$. This may reflect in part that the ratio of atmospheric to soil CO_2 concentration is greater than zero, and that soil water at the evaporating front can be slightly enriched compared to bulk soil water. We conclude that for the present purposes, δ_r is well represented by δ_p , equilibrated with CO_2 , minus 7.6%.

Using the last parameterization of δ_e , 10° latitudinal averages of Δ_A and δ_r (weighted by GPP) were calculated and are presented, with their sum, in Fig. 4, together with δ_o and δ_a . The significance of the sum lies in the future use of this model to predict local δ_a . This will require the linking of surface isotope exchange with an atmospheric mixing model. The local values of δ_a will be smoothed compared with the local isotope 'equilibrium values' (δ_{eq}) that would occur if there were no mixing. For oceans, $\delta_{eq} = \delta_o$, and for the terrestrial biosphere, $\delta_{eq} = (\Delta_A + \delta_r)$. Surface isotope effects will be dominated south of 20°S by oceans, and we note that δ_o increases with latitude as a consequence of lower sea-surface temperatures². Surface effects will be increasingly dominated by the terrestrial biosphere up to 65°N . At this scale, Δ_A and δ_r tend to mirror each other over much of the globe (when averaged annually over a 10°

band), but there is considerable disequilibrium at high northern latitudes. $(\Delta_A + \delta_r)$ decreases markedly above 45°N to a minimum in the $70^\circ\text{--}80^\circ\text{N}$ band (Fig. 4). This is largely a consequence of more negative δ_s (Fig. 2). It is consistent with, and provides an explanation for, measurements of mixed atmosphere δ_a (ref. 1), in which there was a meridional gradient (with strong depletion at 71°N) that required an enormous isotope exchange flux to maintain.

For any particular biome there is no unique relationship between Δ_A and δ_r , and the seasonal cycle in δ_a at most locations should not therefore have a unique relationship with $p\text{CO}_2$. This is in contrast to $^{13}\text{CO}_2$ and is consistent with variability in diurnal³⁹, seasonal^{32,40} and annual¹ δ_a . Nevertheless, in regions south of 60°N , Δ_A is sometimes close to $(\delta_a - \delta_r)$ and in such conditions a plot of δ_a versus $1/C_a$ yields δ_r at $1/C_a = 0$ (ref. 39). For air over Switzerland, a mean value of -16.4% (PDB- CO_2) was calculated³². Our model gives an average δ_r of -14.7% and an average Δ_A of 19.2% for that region.

The most important conclusion of this work is that, not only do terrestrial fluxes have a different effect on δ_a from oceanic fluxes, but different biomes also have markedly different discriminations (Fig. 3). Accurate measurements of rates of change in δ_a in a comprehensive global network together with appropriate meteorological data should therefore help to resolve the present conflicts about the relative role of the oceans and the terrestrial biosphere in the net uptake of CO_2 (equation (3)). They should also allow the identification of components of the terrestrial biota currently acting as net carbon sinks. □

Received 10 November 1992; accepted 22 March 1993.

1. Francey, R. J. & Tans, P. P. *Nature* **327**, 495–497 (1987).
2. Craig, H. & Gordon, L. I. in *Proc. Conf. Stable Isotopes in Oceanographic Studies and Paleotemperatures* 9–130 (Laboratory of Geology and Nuclear Science, Pisa, 1965).
3. Farquhar, G. D., Hubick, K. T., Condon, A. G. & Richards, R. A. in *Stable Isotopes in Ecological Research* (eds Rundel, P. W., Ehleringer, J. R. & Nagy, K. A.) 21–40 (Springer, New York, 1989).
4. Bottinga, Y. & Craig, H. *Earth planet. Sci. Lett.* **55**, 285–295 (1969).
5. Merlivat, L. *J. chem. Phys.* **69**, 2864–2871 (1978).
6. Cowan, I. R. *Adv. Bot. Res.* **4**, 117–228 (1977).
7. O'Neill, J. R., Adami, L. H. & Epstein, S. *J. Res. U.S. geol. Survey* **3**, 623–624 (1975).
8. Farquhar, G. D., Ehleringer, J. R. & Hubick, K. T. *A. Rev. Plant Physiol. Plant molec. Biol.* **40**, 503–537 (1989).
9. Lloyd, J., Syvertsen, J. P., Kriedemann, P. & Farquhar, G. D. *Plant Cell Environ.* **15**, 873–900 (1992).
10. Yakir, D., Berry, J. A., Giles, L. J., Osmond, C. B. & Thomas, R. B. in *Physiological Processes from the Leaf to the Globe* (eds Ehleringer, J. R. & Field, C. B.) 323–338 (Academic, San Diego, 1993).
11. Yurtsever, Y. & Gat, J. R. in *Stable Isotope Hydrology: Deuterium and Oxygen-18 in the Water Cycle* (eds Gat, J. R. & Gonfiantini, R.) 103–142 (IAEA, Vienna, 1981).
12. Hestenberg, R. & Siegenthaler, U. *Tellus* **B43**, 197–205 (1991).
13. International Atomic Energy Agency, *Environmental Isotope Data No. 1: World Survey of Isotope Concentration in Precipitation* Vols 7–9 (1983–1988).
14. Oort, A. *Natn. Oceanic Atmos. Admin. Prof. Pap.* **14** (1983).
15. Shea, D. J. *Tech. Note natn. Cent. atmos. Res. TN-269 + STR* (NCAR, Boulder, 1986).
16. Cogley, J. G. *Tech. Note 85-4* (Trent Univ., Peterborough, Ontario, 1985).
17. Jouzel, J. et al. *J. geophys. Res.* **96**, 7495–7507 (1991).
18. Lloyd, J. *Austr. J. Plant Physiol.* **18**, 649–660 (1991).
19. Cowan, I. R. & Farquhar, G. D. *Symp. Soc. exp. Biol.* **31**, 471–505 (1977).
20. Henderson-Sellers, A., Wilson, M. F., Thomas, G. & Dickinson, R. E. *Natn. Cent. for atmos. Res. Tech. Note-272 + STR* (1986).
21. Lieth, H. in *Primary Productivity of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) 237–263 (Springer, Berlin, 1975).
22. Taylor, J. A., Brasseur, G. P., Zimmerman, P. R. & Cicerone, R. J. *J. geophys. Res.* **96**, 3013–3044 (1991).

23. Poorter, H., Remkes, C. & Lambers, H. *Pl. Physiol.* **94**, 621–627 (1990).
24. Masle, J., Farquhar, G. D. & Gifford, R. M. *Austr. J. Plant Physiol.* **17**, 465–487 (1990).
25. Jarvis, P. & Leverenz, J. W. in *Encyclopaedia of Plant Physiology* Vol. 12D (eds Lange, O. L., Nobel, P. S., Osmond, C. B. & Zeiger, H.) 233–280 (Springer, Berlin, 1983).
26. Medina, E. & Klinge, H. in *Encyclopaedia of Plant Physiology* Vol. 12D (eds Lange, O. L., Nobel, P. S., Osmond, C. B. & Zeiger, H.) 281–304 (Springer, Berlin, 1983).
27. Poorter, H. in *Causes and Consequences of variations in Growth Rate and Productivity of Higher Plants* (eds Lambers, H., Cambridge, M. L., Konings, H. & Pons, T. L.) 49–68 (SPB Academic, The Hague, 1989).
28. Zundel, G., Miekeley, W., Grisi, B. M. & Förstel, H. *Radiat. Envir. Biophys.* **15**, 203–212 (1978).
29. Barnes, C. J. & Allison, G. B. *J. Hydrol.* **100**, 143–176 (1988).
30. Bariac, T., Rambal, S., Jusserand, C. & Berger, A. *Agric. For. Meteor.* **48**, 263–283 (1989).
31. Bariac, T., Jusserand, C. & Mariotti, A. *Geochim. cosmochim. Acta* **54**, 413–424 (1990).
32. Friedli, H., Siegenthaler, U., Rauber, D. & Oeschger, H. *Tellus* **B39**, 80–88 (1987).
33. Tans, P. P. *Tellus* **B32**, 464–469 (1980).
34. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. *Climate Change: The IPCC Scientific Assessment* (Cambridge University Press, UK, 1990).
35. Levitus, S. *Climatological Atlas of the World Ocean*. (Natn. Oceanic Atmosp. Admin. Prof. Pap. **13**, 1981).
36. Friedman, I. & O'Neil, J. R. in *US geol. Surv. Prof. Paper*. 440-KK (1977).
37. O'Neil, J. R. & Adami, L. H. *J. phys. Chem.* **73**, 1553–1558 (1969).
38. Allison, G. B., Colin-Kaczala, C., Filly, A. & Fontes, J. C. H. *J. Hydrol.* **95**, 131–141 (1987).
39. Keeling, C. D. *Geochim. cosmochim. Acta* **24**, 277–298 (1961).
40. Mook, W. G., Koopmans, M., Carter, A. F. & Keeling, C. D. *J. geophys. Res.* **88**, 10915–10933 (1983).

ACKNOWLEDGEMENTS. We thank I. R. Cowan, C. B. Osmond, J. A. Berry, J. R. Evans, J. E. Robertson and S. von Caemmerer for comments, R. Francey, A. Chivas and J. Olley for isotope standards, R. Koster for simulations of $\delta^{18}\text{O}$ of water vapour in relation to precipitation and I. Watterson for humidity fields.

A new species of living bovid from Vietnam

Vu Van Dung, Pham Mong Giao, Nguyen Ngoc Chinh, Do Tuoc, Peter Arcander & John MacKinnon[†]

Forest Inventory and Planning Institute, Ministry of Forestry, Hanoi, Vietnam

* Asian Bureau for Conservation, 18/E Capital Building, 175–191 Lockhart Road, Wanchai, Hong Kong

IN May 1992 a joint survey by the Ministry of Forestry and World Wide Fund for Nature of the Vu Quang Nature Reserve, Ha tinh province, found three sets of long straight horns of a new bovid (Mammalia, Artiodactyla) in hunters' houses¹. None of the specimens had dentition. On four follow-up visits by Vietnamese scientists new specimens were discovered and surveys of forests in neighbouring Nghe an province revealed more localities and some partial specimens. In all, we have examined more than 20 specimens. Three have complete upper skulls and dentitions, two have lower jaws and dentitions. Three complete skins have been collected. The specimens are distinct in appearance, morphology and DNA sequence and cannot be ascribed to any known genus. Only two bovid genera are known from this part of Asia, *Bos* and *Naemorhedus* = *Capricornis*^{2,3}. A new genus and species are therefore described. Such a discovery is of great significance. It has been more than 50 years since any comparable find of a large mammal species has been made; the last being the kouprey *Bos* = *Novibos sauveli*, another Indochinese bovid (Urbain, 1937). Moreover, the bovids (cattle, goats and antelopes) are a mammal family of great value to mankind. Many species have proven or potential value for domestication or cross-breeding. A three-month field study is planned to observe the living animal.

FORMAL DESCRIPTION

Family Bovidae
Subfamily Bovinae
Tribe ? Boselaphini

Genus *Pseudoryx*, gen. nov.

Diagnosis. *Pseudoryx* differs significantly from all described genera in appearance, morphology, cranial and dental features and DNA. The long, smooth, almost straight, slender horns, elongated premolars, large face gland and distinctive colour

pattern are diagnostic.

Description. See under species description.

Etymology. The name reflects the superficial similarity to *Oryx* in having long straight horns slightly recurved in profile, with bold black and white facial markings, while clarifying that the animal is not closely related to *Oryx*.

Type species. *P. nghetinhensis*, sp. nov.

Pseudoryx nghetinhensis sp. nov.

Diagnosis. The only species of the genus. Diagnosis as for genus.

Description. The total body weight of the adult is estimated at about 100 kg. Total length from nose to anus is about 1.5 m. Height at the shoulder is about 80–90 cm. Length from spine to front foot across preserved skin is 96 cm. Tail is short, about 13 cm of bone with fluffy black tassel. Ear length is quite short at about 10 cm. Skull length varies between 32 and 36.5 cm.

The skull is highly bridged in the nasal area. The horns are long, almost straight, smooth, almost parallel in females, and only moderately diverging in males. They are almost circular in cross-section, with horn cores extending close to the tip. Horn length varies from 32 to 52 cm, with a mean (18) = 41 cm. Width between tips varies from 7.5 to 20 cm, with mean (17) = 13.3 cm. Both length and distance between the tips show a bimodal distribution, with inferred males having longer more divergent horns than females. The insertion of the horns is much wider than for goat-antelopes (such as the serow *Naemorhedus sumatraensis*; Fig. 3b), similar to the mountain anoa *Bubalus quarlesi*, but narrow for cattle. Internal width between horns basally varies from 3.0 to 4.0, with a mean (6) = 3.7 cm. The outer width across the horns basally varies from 10 to 12, with a mean (6) = 10.5 cm. The basal 7 cm shows narrow annuli, but

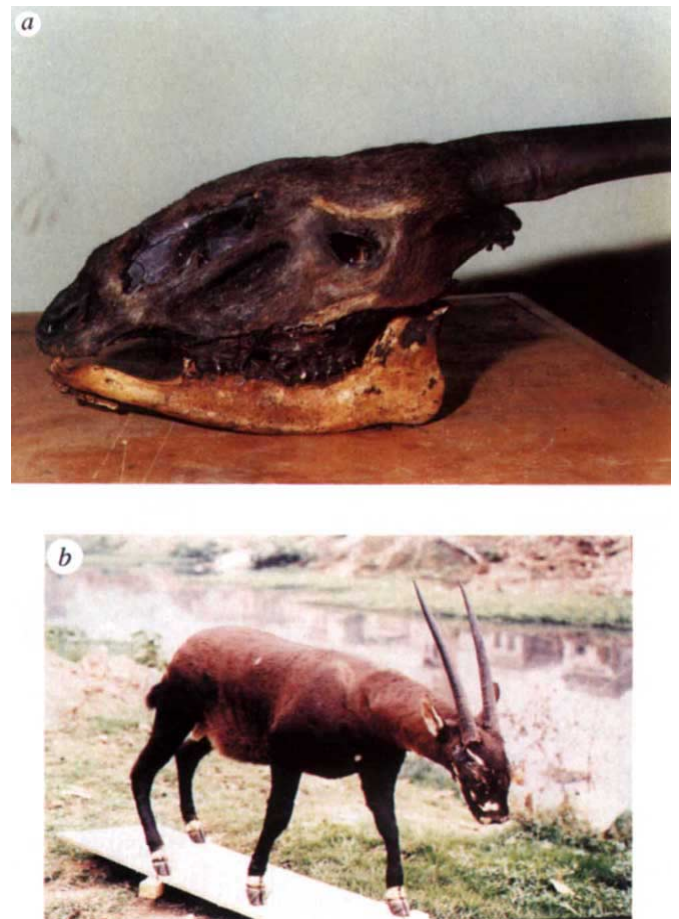


FIG. 1 Photographs of Vu Quang bovid: a, type specimen FIPI/MVQ001; b, stuffed skin of another individual.

[†] To whom correspondence should be addressed.

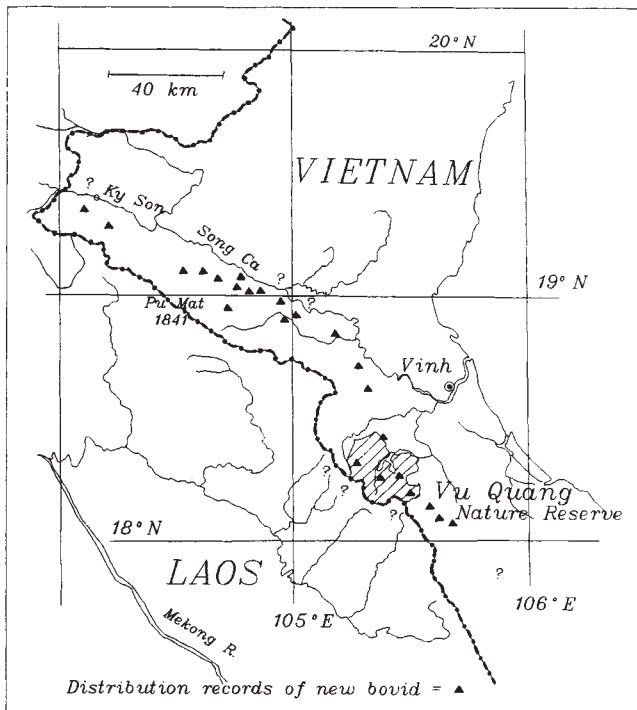
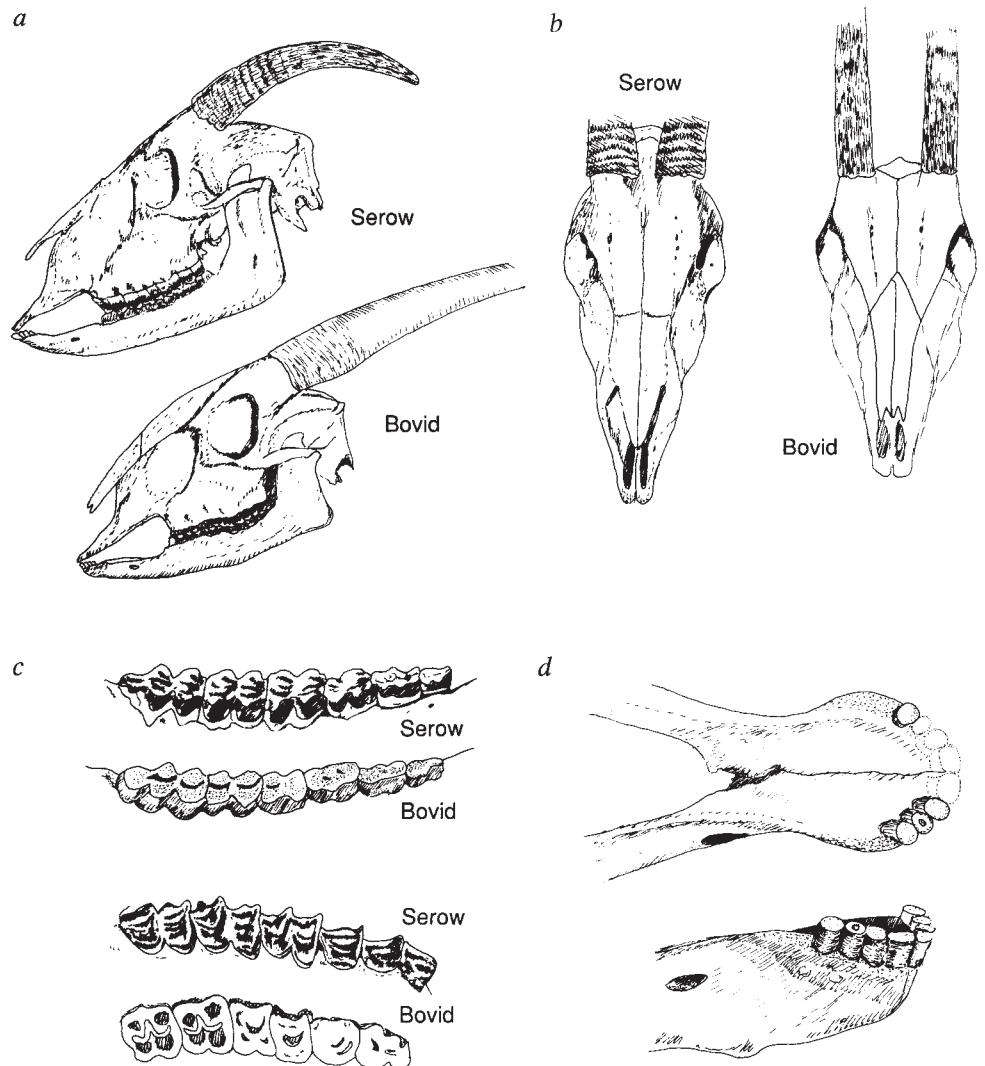


FIG. 2 Distribution of Vu Quang bovid. Definitive records are now known from over 20 localities, all on the forest edge along the cordillera along the Laos-Vietnam border between 104°5'E by 19°25'N and 105°50' by 18°05'N. All specimens are from the Vietnam (wetter) side of the range, but hunters report that the animal does also exist on the Lao side. All definitive records are to the south side of the Song Ca river.

FIG. 3 Cranial details of new bovid. *a*, Full skull (composite reconstruction); *b*, frontal view; and *c*, molar and premolar toothrows of Vu Quang bovid compared with serow. Serow is the only bovid of comparable size in Southeast Asia.



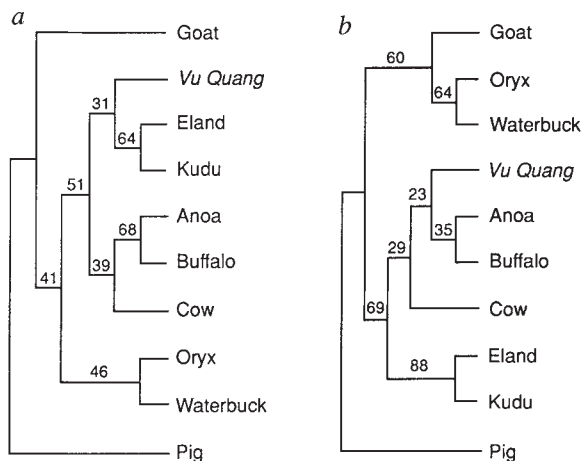


FIG. 4 Phylogenetic relationship of the Vu Quang bovid in relation to representatives of some major bovid groups. Shown is a maximum parsimony bootstrap tree⁵ based on 249-base-pair mitochondrial DNA encoding cytochrome *b*, with the pig (*Sus scrofa*) used as outgroup. Numbers above branches refer to the frequency (in per cent) of finding a group to be monophyletic for 100 bootstrapped samples of the sequences. All substitutions are used in *a* (170 steps, 95 segregating and 66 informative sites). Because of the age of these nodes, transitions in the third codon position are likely to be affected by saturation, and tree *b* shows the results when these are omitted (66 steps, 27 informative sites). In both trees, the Vu Quang bovid clusters with but does not appear especially close to cow (*Bos taurus*), buffalo (*Syncerus caffer*), anoa (*Bubalus depressicornis*), eland (*Tragelaphus oryx*) and kudu (*Tragelaphus imberbis*). The low bootstrap numbers indicate that more sequence data are needed to fully resolve the branching order. But these DNA data show affiliation with the Bovinae as opposed to other bovid groups. DNA extractions, amplifications using the polymerase chain reaction, and sequencing all followed standard procedures^{6,7}. The primers HL14841 and HH15149 (ref. 8) were used for amplifications as well as for sequencing. Note: The sequences from pig, cow and goat (*Capra hircus*) are published⁹; oryx (*Oryx gazella*), waterbuck (*Kobus ellipsiprymnus*), buffalo, kudu and eland sequences were obtained under a joint project between the National Museums of Kenya and the Institute of Population Biology, University of Copenhagen, Denmark; anoa DNA was kindly donated by the Conservation Genetics Group, Institute of Zoology, London, UK; Vu Quang DNA is extracted from a smoke-dried skin sample¹⁰. Sequences are deposited in GenBank under the accession numbers: L13792, anoa; L13793, eland; L13794, kudu; L13795, oryx; L13796, buffalo; L13797, *Pseudoryx*; L13798, waterbuck.

the main shaft is rubbed smooth.

Dental formula is $\begin{matrix} 0 & 0 & 3 & 3 \\ 3 & 1 & 3 & 3 \end{matrix} \times 2 = 32$

Available dentitions of *Pseudoryx* are in late occlusal wear. The lower front teeth appear less spatulate than in other genera (Fig. 3d). The second and third incisors and canine of the type specimen are peglike and almost vertical in alignment rather than splayed forward. The premolars are elongated (Fig. 3e). The proportion of molar and premolar tooth-row length accounted for by the premolars is 44% (lower) and 46% (upper), compared with 39% and 36% for the serow.

The face and general body colour varies from dark brown to rich reddish brown. There is a blackish brown narrow stripe down the middle of the back onto the short tail. Both sexes have whitish-to-buff stripes above and below the eye, and white patterns on the side of the face, chin and throat. All skins show a whitish stripe on the outer rump separating the brown back from blackish legs and show whitish socklet markings with black frontal division. The hair on the upperparts is fine, short (about 2.4 cm) and glossy, whereas on the underparts it is longer (3.5 cm) and more fluffy. The anal area and inner flanks are whitish, as is the scrotum of the male skin. The ear is brown behind and whitish inside, with short hair, except for a tuft of long whitish hair on the upper innerside. The hair pattern results in whorls on the centre of the nose, the sides of the neck and middle of the shoulders. The animal has no dorsal crest of long hair as in serow and nilgai *Boselaphus tragocamelus*.

The hooves are small and dainty with short, blunt toes about 4 cm high and 6 cm long. The 'dew toes' are vestigial calluses, much more reduced than in goat-antelopes or cattle. There are large face glands just in front of the eye as in serows, duikers (*Cephalophini*) and some other bovids.

Etymology. The old Vietnamese province Nghe tinh includes the two new provinces Nghe an and Ha tinh, which precisely define the known distribution of the species.

Holotype. Forest Inventory and Planning Institute (FIPI) Museum, Hanoi. Specimen number FIPI/MVQ001, adult male, skin of head and almost complete skull with lower mandibles, from Vu Quang nature reserve, Vietnam 105°25'E by 18°15'N, collected by FIPI team, April 1992 (see Figs 1 and 3d).

Distribution. Ha tinh and Nghe an provinces, Vietnam (Fig. 2).

Currently known range totals about 4,000 km² and constitutes the only extensive pristine forest in northern Vietnam. Montane forests are dominated by conifers with *Fokienia hodgsoni*, the main species. Lower forests are rich, evergreen, mixed-broad-leaf, dominated by the dipterocarp *Hopea mollissima*. The new bovid appears to use all forest levels in differing seasons, even coming into the secondary lowland forests along large rivers at about 200 m. The highest peaks in the range are over 2,000 m above sea level. From the size of the distribution area and the frequency with which hunters claim to trap the animals, we estimate that a few hundred of the new bovid survive. Following this discovery, the Vietnamese Government has increased the size of the Vu Quang nature reserve from 16,000 ha to 60,000 ha and has plans for another two reserves.

Preliminary DNA analysis indicates that the new species clusters with oxen (Bovinae) rather than other bovids (S. J. O'Brien, personal communication; P. A. Arcander and S. J. O'Brien, manuscript in preparation; also see Fig. 4). The DNA analysis does not yet indicate clearly to which tribe of Bovinae the species belong. The species appears to be a rather deep branch of the subfamily. It may merit creation of a new tribe. It exhibits many anatomical characters listed as primitive for bovids⁴, and our placement in tribe Boselaphini (one tribe of subfamily Bovinae) is provisional and based on morphological features. Only this tribe shows the combination of primitive features of face glands with the pattern of black and white markings on face, neck, feet and rump (c.f. nilgai). □

Received 7 April 1993; accepted 11 May 1993.

- MacKinnon, J. & Vu Van Dung. *Technical Report of MOF/WWF* (1992).
- Corbet, G. B. & Hill, J. E. *The Mammals of the Indomalayan Region* (Oxford Univ. Press, 1992).
- Lekagul, B. & McNeely, J. A. *Mammals of Thailand* (Assoc. Conservation Wildlife, Bangkok, 1977).
- Gentry, A. W. *Mamm. Rev.* **22**, 1–32 (1992).
- Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony* Version 3.0. (Illinois Hist. Surv., Campaign, 1990).
- Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. *PCR Protocols* (Academic, San Diego, 1990).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Kocker, T. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6196–6200 (1989).
- Irwin, D. M., Kocker, T. D. & Wilson, A. C. *J. molec. Evol.* **43**, 128–144 (1991).
- Smith, E. F. G., Arcander, P., Fjeldsa, J. & Amir, O. G. *Ibis* **133**, 227–235 (1991).

ACKNOWLEDGEMENTS. We thank A. Gentry, C. Groves, P. Grubb, S. Stuart and G. Corbet for their comments on photos of the specimens and drafts of this paper.

Naturally occurring antibodies devoid of light chains

C. Hamers-Casterman, T. Atarhouch, S. Muyldermans, G. Robinson*, C. Hamers, E. Bajana Songa, N. Bendahman & R. Hamers†

Vrije Universiteit Brussel, Instituut voor Moleculaire Biologie, Paardenstraat 65, B-1640 Sint Genesius Rode, Belgium

* Department of Biochemistry, Trinity College, Dublin 2, Ireland

RANDOM association of VL and VH repertoires contributes considerably to antibody diversity¹. The diversity and the affinity are then increased by hypermutation in B cells located in germinal centres². Except in the case of 'heavy chain' disease³, naturally occurring heavy-chain antibodies have not been described, although antigen binding has been demonstrated for separated heavy chains⁴ or cloned VH domains⁵. Here we investigate the presence of considerable amounts of IgG-like material of M_r 100K in the serum of the camel (*Camelus dromedarius*)⁶. These molecules are composed of heavy-chain dimers and are devoid of light chains, but nevertheless have an extensive antigen-binding repertoire, a finding that calls into question the role of light chains in the camel. Camel heavy-chain IgGs lack CH1, which in one IgG class might be structurally replaced by an extended hinge. Heavy-chain IgGs are a feature of all camelids. These findings open new perspectives in the engineering of antibodies.

By a combination of affinity chromatography on protein A and protein G, three quantitatively important fractions corresponding to subclasses of IgG can be isolated from the serum of camels (Fig. 1A, lanes c–f).

One fraction (IgG₁) contains molecules of M_r 170K (Fig. 1B, lane 2), which upon reduction yield 50K heavy chains and large 30K light chains (Fig. 1C, lane 2). The two other immunoglobulin fractions contain molecules of ~100K (Fig. 1B, lanes 1 and 3), which upon reduction yield only heavy chains of, respectively, 46K (IgG₂ fraction binding only to protein A) (Fig. 1C, lane 3) and 43K (IgG₃ fraction binding to protein A and protein G) (Fig. 1C, lane 1). These two IgG classes appear to lack the light chain completely.

To exclude the possibility that the light chains were only weakly associated with the heavy chains and lost during our selective purification, whole serum was size-fractionated by gel filtration. Coomassie blue staining of unreduced fractions revealed the sequential elution of the 170K IgG₁ followed by the incompletely resolved isotypes IgG₂ and IgG₃ (90K) (Fig. 1D, upper inset). Immunostaining of the same fractions after reduction confirmed that the light chains were present only in the 50K heavy-chain-containing fractions (Fig. 1D, lower inset).

A comparative study of old world camelids (*Camelus bactrianus* and *Camelus dromedarius*) and new world camelids (*Lama pacos*, *Lama glama* and *Lama vicugna*) showed that heavy-chain immunoglobulins are abundant in the sera of all species examined (data not shown) and total up to 75 per cent of the molecules binding to protein A.

The abundance of heavy-chain immunoglobulins in the serum of camelids raises the question as to whether they bear an extensive antigen-binding repertoire. This question could be answered by examining the IgG₁, IgG₂ and IgG₃ fractions from the serum of camels (*Camelus dromedarius*) with a high antitypanosome titre⁷. In radio-immunoprecipitation, purified fractions of IgG₁, IgG₂ and IgG₃ derived from infected camels were shown to bind a large number of antigens present in a [³⁵S]methionine-labelled trypanosome lysate (Fig. 2a), indicating an extensive repertoire complexity for the three IgG classes. Conversely, in blotting experiments, [³⁵S]methionine-labelled trypanosome lysate binds to SDS-PAGE-separated IgG₁, IgG₂

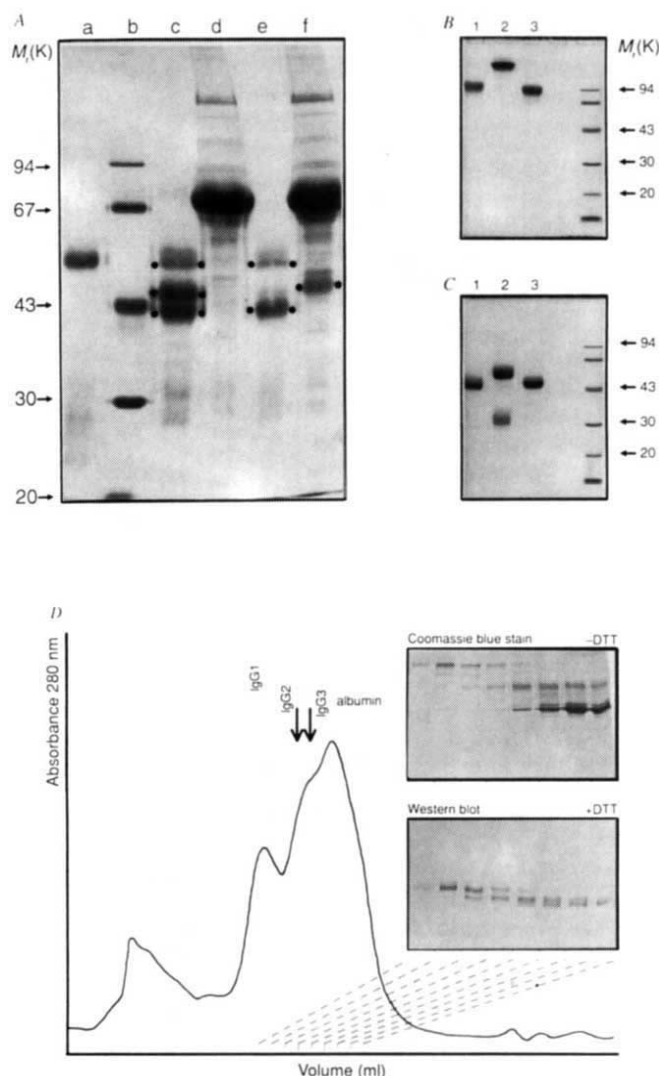


FIG. 1 Characterization and purification of camel IgG classes on protein A, protein G and gel filtration. **A**, The fraction of *C. dromedarius* serum adsorbed on protein A on reducing SDS-PAGE shows three heavy-chain components of 50, 46 and 43K (bands between dots), which are absent in the non-adsorbed fraction (lane d), and light-chain components of ~30K (lane c) that are considerably larger than rabbit light chain (lane a, rabbit IgG). The fractions adsorbed on protein G (lane e) lack the 46K heavy chain which remains in the non-adsorbed fraction (lane f). Lane b contains a size marker. **B** and **C**, By differential adsorption and elution on protein G and protein A, the IgG fractions containing 43K (lane 1), 46K (lane 3) and 50K (lanes 2) heavy chains were purified and analysed on SDS-PAGE in the absence (**B**) or presence (**C**) of dithiothreitol (DTT). **D**, Camel whole serum (0.1 ml) was fractionated by gel filtration on a Superdex-200 column using 150 mM NaCl, 50 mM sodium phosphate buffer, pH 7.0, as eluent. Affinity-purified IgG₂ and IgG₃ elute at the positions indicated by arrows. Fractions of interest were analysed further by SDS-PAGE with or without prior reduction. The protein contents as visualized by Coomassie blue (without reduction; upper inset) are compared with the immunoglobulins from the same fractions (after reduction with DTT, lower inset) as revealed by western blotting with a rabbit anti-camel IgG (lower inset).

METHODS. *C. dromedarius* serum (5 ml) is adsorbed onto a 5-ml protein G-Sepharose (Pharmacia) column and washed with 20 mM phosphate buffer, pH 7.0. Upon elution with 0.15 M NaCl, 0.58% acetic acid (pH 3.5), IgG₃ of 100K is eluted, which upon reduction yields heavy chains of 43K (lanes 1 in **B** and **C**). IgG₁ of 170K can subsequently be eluted with pH 2.7 buffer (0.1 M glycine-HCl). This fraction upon reduction yields a 50K heavy-chain and a broad light-chain band (lane 2 in **C**). The fraction not adsorbed on protein G is run on a 5-ml protein A-Sepharose column. After washing and elution with 0.15 M NaCl, 0.58% acetic acid (pH 4.5), IgG₂ of 100K is obtained, which consists solely of 46K heavy chains (lane 3 in **C**).

† To whom correspondence should be addressed.

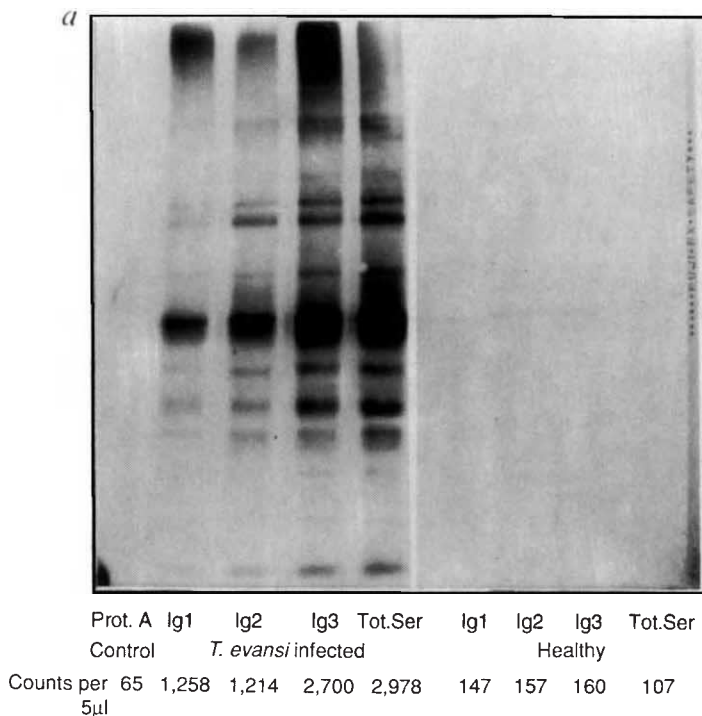
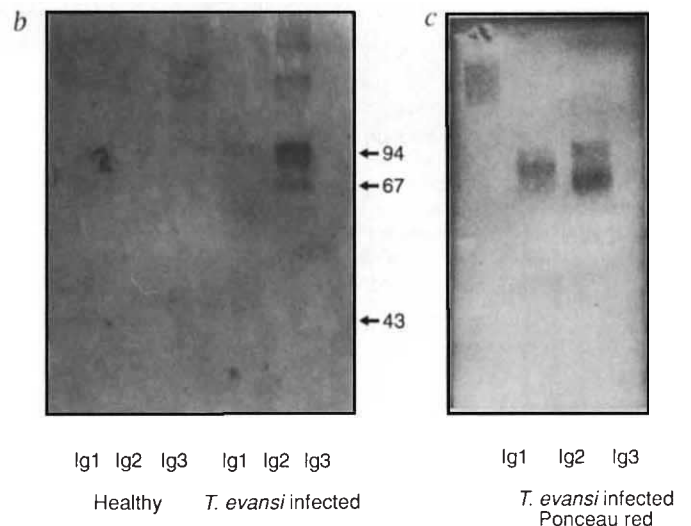


FIG. 2 Repertoire complexity and antigen-binding capacity of camel IgG₁, IgG₂ and IgG₃ analysed by *a*, radioimmunoprecipitation, or *b* and *c*, western blotting. *a*, Serum or purified IgG fractions from healthy or *Trypanosoma evansi*-infected *C. dromedarius* (CATT titre 1/160 (ref. 7)) were incubated with labelled trypanosome lysate, recovered with protein A-Sepharose and analysed by SDS-PAGE. The relative counts recovered are inscribed below each lane. No trypanosome proteins bind to the protein A or to the healthy camel immunoglobulins. *b*, 20 µg of IgG₁, IgG₂ and IgG₃ from healthy and trypanosome-infected animals were separated by SDS-PAGE without prior reduction or heating. The electroblotted proteins were incubated with the labelled trypanosome lysate. The IgG₂ shows a single antigen-binding component corresponding to the heavy-chain immunoglobulin, whereas the IgG₃



fraction appears in addition to contain two larger antigen-binding components barely detectable by Ponceau red staining (*c*). These could possibly be immunoglobulin classes copurified as immunocomplexes present in the serum of the infected animals.

METHODS. [³⁵S]methionine-labelled *Trypanosoma evansi* lysate (500,000 counts)²² was incubated (4 °C, 1 h) with 10 µl serum or, 20 µg IgG₁, IgG₂ or IgG₃ in 200 µl of 0.4 M NaCl, 10 mM EDTA, 10 mM Tris, pH 8.3, containing 0.1 M *N*-α-tosyl-L-lysine chloromethyl ketone (TLCK). Protein A-Sepharose (10 mg) suspended in 200 µl of the same buffer was added (4 °C, 1 h). After washing and centrifugation, each pellet was resuspended in 75 µl SDS-PAGE sample solution containing DTT and heated for 3 min at 100 °C. After centrifugation, 5 µl supernatant was saved for radioactivity counting and the remainder analysed by SDS-PAGE and fluorography. The nitrocellulose filter of the western blot of purified fractions IgG₁, IgG₂ and IgG₃ was stained with Ponceau red (*c*) or incubated with 1% ovalbumin in TST buffer (Tris 10 mM, NaCl 150 mM, Tween 0.05%) (*b*). The membrane was extensively washed with TST buffer and incubated for 2 h with ³⁵S-labelled trypanosome antigen. To avoid nonspecific binding, labelled trypanosome antigen lysate was filtered (45 µm) and incubated with healthy camel immunoglobulin and ovalbumin adsorbed on a nitrocellulose membrane.

```

10      20      40
EVQLVESGGG LVQPQGSRLRL SCAASG CDR1 VNRQA PGKLEWVS CDR2
GG SVQGGGSLRL SCAISG CDR1 WFREG PGKEREKIA CDR2
GG SVQAGGSLRL SCASSS CDR1 WYRQA PGKEREVVS CDR2

70      80      90      110
RFTIS RDNSKNTLYL QMNSLRAEDTAVY YCAR CDR3 WGQGTQVT VSS
RFTIS QDSTLKTMYL LMNKLKPEDTGY YCAA CDR3 WGQGTQVT VSS
RFTIS QDSAKNTVYL QMNSLKPEDTAMY YCKI CDR3 WGQGTQVT VSS

```

```

Camel VH      Hinge      CH2
WGQGTQVT VSS — GTNEVCKCKPCP APELPGG PSVFVFP
WGQGTQVT VSS — EPKIPQPPKPPQPP
                  QPQPPQPP
                  KPEPCKTCPCPCP APELLGG PSVFIFP

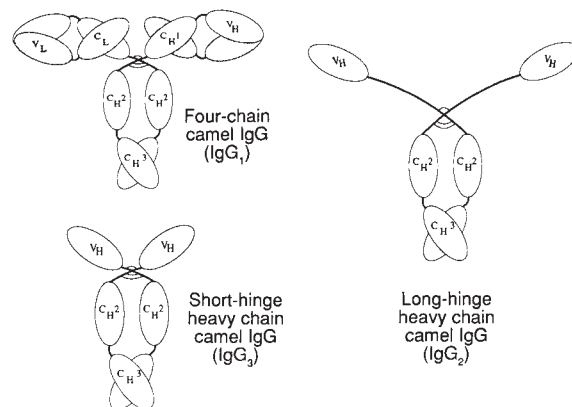
Human CH1      Hinge      CH2
KVDRKV — ELKTPGLDTHTCPCPCP
                  EPKSCDTPPPCPCPCP
                  EPKSCDTPPPCPCPCP APELLGG PSVFLFP
gamma 1 KVDRK — AEPKSCDTHTCPCPCP APELLGG PSVFLFP
Human γ-2 KVKVT — ERKCCVECPPCP APPVAG PSVFLFP
Human γ-4 KVDRV — ESKYPPPCPCPCP APELLGG PSVFLFP

```

FIG. 3 Amino-acid sequences of the V_H framework, and hinge/CH₂ of *Camelus dromedarius* heavy-chain immunoglobulins, compared to human (italic) V_H framework (subgroup III) and hinges of human IgG¹⁴.

METHODS. Total RNA was isolated from a dromedary spleen²³. mRNA was purified with oligo(T) paramagnetic beads (PolyAtract, Promega). 1 µg mRNA was used for preparing double-stranded cDNA²³ after an oligo(dT) priming using enzymes from Boehringer Mannheim. 5 ng cDNA was amplified by PCR in a 100-µl reaction mixture (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 15 mM MgCl₂, 0.01% (w/v) gelatine, 200 µM of each dNTP). 25 pmol of each oligonucleotide of mouse V_H²⁴, containing an *Xho*I site, and 5'-CGCCAT-CAAGGTACCAAGTGA-3' were used as primers. The 3'-end primer was deduced from partial sequences corresponding to γ-chain amino acids 296 to 288 (T. Atarhouch, C. H.-C. and G.R., unpublished results) in which one mismatch was introduced to create a *Kpn*I restriction site. After a round of denaturing annealing (94 °C for 5 min and 54 °C for 5 min), 2 U of *Taq* DNA polymerase were added to the reaction mixture before 35 cycles of amplification⁵. PCR products were purified by phenol-chloroform extraction followed by HPLC (Genpak-fax column, Waters) and finally by MERMAID (BIO101). After purification, amplified cDNA was digested with *Xho*I and *Kpn*I, and ligated into pBluescript. Clones were sequenced by dideoxy chain termination²⁵. Sequences were then translated so that they could be assigned to well defined domains of the immunoglobulin molecule¹⁴.

FIG. 4 Schematic representation of the structural organization of the camel immunoglobulins (adapted from ref. 26). On the basis of size, the IgG₁ fraction probably has the normal antibody assembly of two light and two heavy chains. IgG₃ would have a hinge comparable in size to the human IgG₁, IgG₂ and IgG₄. The two antigen-binding sites are much closer to each other as camel IgG lacks the CH1 domain. In camel IgG₂ the long hinge, which is formed of Pro-X repeats (X=Glu, Gln or Lys), probably adopts a rigid structure^{19,20}. This long hinge could therefore substitute for the CH1 domain and bring the two antigen-binding sites of IgG₂ into normal positions.



and IgG₃ obtained from infected animals (Fig. 2b, c). These findings indicate that the heavy chains alone can generate an extensive repertoire and question the obligatory contribution of the light chain to the useful antibody repertoire in the camelids.

The camelid $\gamma 2$ and $\gamma 3$ chains are considerably shorter than the normal mammalian γ or camel $\gamma 1$ chains. This would suggest that, as in the case of 'heavy chain' disease³, deletions have occurred in the CH1 protein domain^{8,9}. To address this question, complementary DNA was synthesized from camel spleen messenger RNA and the sequences between the 5' end of the VH and the CH2 were amplified by polymerase chain reaction (PCR), and cloned. Seventeen clones presenting a different VH sequence were isolated and sequenced. Their most striking feature was a complete lack of the CH1 domain, the last framework (FR4) residues of the VH region being immediately followed by the hinge (Fig. 3, lower part). The absence of the CH1 domain clarifies two important dilemmas. First, immunoglobulin heavy chains are normally not secreted unless the heavy-chain chaperoning protein or BIP (ref. 10) has been replaced by the L chain¹¹, or alternatively the CH1 domain has been deleted^{3,8,9}. Secondly, isolated heavy chains from mammalian immunoglobulins tend to aggregate, but are only solubilized by light chains^{8,12} which bind to the CH1 and the VH domains¹³.

Fourteen of the seventeen clones were characterized by a short hinge sequence with a length equal to that of human IgG₂ and IgG₄ (ref. 14) (Fig. 3). The other three had a long hinge sequence containing the 'EPK' hinge motif found in human IgG₁ and IgG₃ (ref. 14). They possess the CH2 'APELL/P' motif that is also found in human IgG₁ and IgG₃, and which is associated with mammary transport of bovine IgG₁ (ref. 15). On the basis of their molecular weights, we expect the 'short-hinge' clones to correspond to IgG₃ and the 'long-hinge' clones to IgG₂.

In the short hinge-containing antibody, the extreme distance between the extremities of the VH regions will be of the order of 80 Å, corresponding to twice the size of a single domain of 40 Å ($2 \times V_H$)¹⁶. This could be a severe limitation for agglutination, crosslinking or complement fixation^{17,18}. In the long-hinge-containing immunoglobulin, the absence of CH1 might be compensated by the extremely long hinge itself, composed of a 12-fold repeat of the sequence Pro-X (where X is Gln, Glu or Lys) (Figs 3 and 4). NMR (ref. 19) and molecular modelling²⁰ of Pro-X repeats present in the TonB protein of *Escherichia coli* (in which X is Glu or Lys) and of the membrane procyclin of trypanosomes (X is Asp or Glu) indicate that these repeated sequences function as rigid rod-like spacers with a diameter of 8 Å and a rise of 2.9 Å per residue. Assuming the same geometry, the long hinge would be 70 Å, which compensates for the absence of the CH1 domain.

The binding site of heavy-chain antibodies cannot form the pocket resulting from adjoining light and heavy chain V regions, and the residues of VH that normally interact with VL will be exposed to solvent^{3,5,13}. Leucine at position 45 is conserved in 98% of human and murine VH sequences¹⁴ and is crucial in the VH-VL association¹³; it can be replaced by an arginine (Fig. 3, top section). This substitution is in accordance with both the lost contact with a VL domain and an increased solubility.

Unlike myeloma heavy chains, which result mainly from CH1 deletion in a single antibody-producing cell²¹, the camelid heavy-chain antibodies have emerged in a normal immunological environment and will probably have undergone the selective refinement in specificity and affinity that accompanies B-cell maturation^{1,2}. The obtention of camelid heavy-chain antibodies could therefore be an invaluable asset in the development and engineering of soluble VH domains⁵ or of new immunological molecules for diagnostic, therapeutic and biochemical purposes.

Received 20 January; accepted 22 March 1993.

1. Tonegawa, S. *Nature* **302**, 575-581 (1983).
2. Jacob, J., Kelsøe, G., Rajewski, K. & Weiss, U. *Nature* **354**, 389-392 (1991).
3. Fleischman, J. B., Pain, R. H. & Porter, R. R. *Arch. Biochem. Biophys. Suppl.* **1**, 174-180 (1962).
4. Utsumi, S. & Karush, F. *Biochemistry* **3**, 1329-1338 (1964).
5. Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. *Nature* **341**, 544-546 (1989).
6. Ungar-Waron, H., Eliase, E., Gluckman, A. & Trainin, Z. *Isr. J. vet. Med.* **43**, 198-203 (1987).
7. Baijaya Songa, E. & Hamers, R. *Ann. Soc. Belge Méd. Trop.* **68**, 233-240 (1988).
8. Seligmann, M., Milhaes, E., Preud'homme, J.-L., Danon, F. & Brouet, J.-C. *Immun. Rev.* **48**, 145-167 (1979).
9. Trautwein, A., Schneider, J., Kiefer, H. & Karjalainen, K. *Nature* **339**, 68-70 (1989).
10. Hendershot, L. M., Bole, D., Köhler, G. & Kearney, J. F. *J. Cell Biol.* **104**, 761-767 (1987).
11. Hendershot, L. M. *J. Cell Biol.* **111**, 829-837 (1990).
12. Roholt, O., Onoue, K. & Pressman, D. *Proc. natn. Acad. Sci. U.S.A.* **51**, 173-178 (1964).
13. Chothia, C., Novotny, J., Brucoleri, R. & Karplus, M. *J. molec. Biol.* **186**, 651-663 (1985).
14. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. *Sequences of Proteins of Immunological Interest 511* (US Dept of Health and Human Services, US Public Health Service, NIH, Bethesda, 1987).

15. Jackson, T., Morris, B. A. & Sanders, P. G. *Molec. Immun.* **29**, 667-676 (1992).
16. Poljak, R. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **70**, 3305-3310 (1973).
17. Dangel, J. L. *et al. EMBO J.* **7**, 1989-1994 (1988).
18. Schneider, W. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 2509-2513 (1988).
19. Evans, J. S. *et al. FEBS Lett.* **208**, 211-216 (1986).
20. Roditi, I. *et al. J. Cell Biol.* **108**, 737-746 (1989).
21. Dunnick, W., Rabbits, T. H. & Milstein, C. *Nucleic Acids Res.* **8**, 1475-1484 (1980).
22. Bülow, R., Nonnengässer, C. & Overath, P. *Molec. Biochem. Parasitol.* **32**, 85-92 (1989).
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
24. Sastry, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5728-5732 (1989).
25. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
26. Klein, J. *Immunology* (Blackwell Scientific, London, 1990).

ACKNOWLEDGEMENTS. We thank A. Dakkak, P. Sayer and R. Dedeken for assistance in obtaining samples of camelid sera, W. Verheulpen for preparing the figures, and L. Eeckhaudt for typing the manuscript. This research was supported by the Biotechnology programme of the Flemish regional Government and by the ECC programmes 'Science and Technology for Development'.

A point mutation in a *Drosophila* GABA receptor confers insecticide resistance

Richard H. ffrench-Constant, Thomas A. Rocheleau,
Jessica C. Steichen & Alison E. Chalmers*

Department of Entomology, 237 Russell Laboratories, 1630 Linden Drive,
University of Wisconsin, Madison, Wisconsin 53706, USA

*Rhône-Poulenc Ag Company, 2 T.W. Alexander Drive,
Research Triangle Park, North Carolina 27709, USA

VERTEBRATES and invertebrates both have GABA (γ -aminobutyric acid) as a major inhibitory neurotransmitter^{1,2}. GABA_A receptors in vertebrates assemble as heteromultimers to form an integral chloride ion channel³. These receptors are targets for drugs and pesticides⁴ and are also implicated in seizure-related diseases^{5,6}. Picrotoxinin (PTX) and cyclodiene insecticides are GABA_A receptor antagonists which competitively displace each other from the same binding site⁷. Insects⁸ and vertebrates⁹ showing resistance to cyclodienes also show cross-resistance to PTX. Previously, we used a field-isolated *Drosophila* mutant *Rdl* (Resistant to dieldrin)¹⁰ insensitive to PTX and cyclodienes to clone a putative GABA receptor¹¹. Here we report the functional expression and novel pharmacology of this GABA receptor and examine the functionality of a resistance-associated point mutation (alanine to serine) within the second membrane-spanning domain, the region thought to line the chloride ion channel pore. This substitution is found globally in *Drosophila* populations¹². This mutation not only identifies a single amino acid conferring high levels of resistance to the important GABA receptor antagonist PTX but also, by conferring resistance to cyclodienes, may account for over 60% of reported cases of insecticide resistance¹³.

To identify the mutation(s) conferring insensitivity, we amplified and sequenced the complete open reading frame of the *Rdl*^{R-MD} allele from the resistant RNA (Fig. 1). Two amino-acid substitutions were detected within the sequence, one an alanine to serine at position 302 (Ala302 → Ser) and the other a methionine to isoleucine at position 361 (Met361 → Ile). To confirm the presence of these mutations in the *Rdl*^{R-MD} allele, complementary DNA clones derived from resistant RNA and the corresponding regions of the genomic DNA were sequenced. After further sequencing of the appropriate exons in genomic DNA from 14 other insensitive *D. melanogaster* alleles collected from a number of different continents, only the alanine to serine substitution was found to be consistently associated with PTX/cyclodiene insensitivity¹². This mutation is within the second membrane-spanning region (M2) (Fig. 1) near the site conferring charge selectivity in the closely related nicotinic acetylcholine receptor^{14,15}. Here we present an analysis of the

role of this substitution in channel insensitivity to PTX and cyclodiene insecticides using the *Xenopus* oocyte expression system (Fig. 2).

Injection of relatively high concentrations of wild-type (*Rdl*^S) messenger RNA alone (>50 ng per oocyte), synthesized *in vitro* from cDNA 14.1, resulted in the appearance of GABA-evoked currents with a reversal potential of -20 to -25 mV (similar to that previously described for chloride channels¹⁶). The GABA response was dose-dependent, with a threshold varying from 0.5 to 100 μ M GABA, and maximum currents ranging from 50 nA to >5 μ A (oocytes were voltage clamped at -60 mV). Similar currents were elicited by the GABA_A agonist muscimol, but GABA-evoked currents were not blocked by the GABA receptor antagonist bicuculline (Fig. 1a). Bicuculline-insensitive

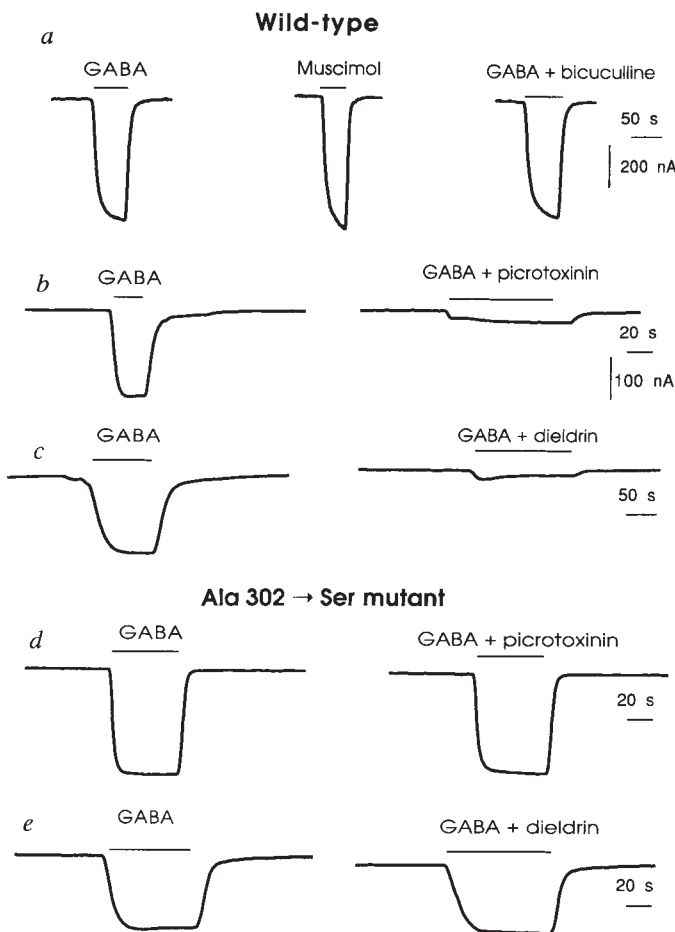


FIG. 2 Currents recorded under voltage-clamp conditions from oocytes injected with (a, b, c) wild-type 14.1 cDNA and (d, e) Ala302 → Ser mutant. a, Responses of a single oocyte to GABA (10 μ M), muscimol (10 μ M) and bicuculline (100 μ M with 10 μ M GABA). b–e, The left column shows currents measured in response to 50 μ M GABA (the rough EC₅₀ concentration required to elicit 50% of the maximum GABA response) and the right column shows the effect on these currents of 1 μ M PTX or 10 μ M dieldrin. The vertical scale in b refers to b–e. Oocytes injected with RNA were voltage clamped at -60 mV with a two-electrode amplifier. Currents were recorded in 96 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂ and 5 mM HEPES. Recordings were made 1–5 days after injection. Picrotoxinin, GABA, muscimol hydrobromide and (-)-bicuculline methiodide were dissolved in saline, dieldrin was dissolved in DMSO then diluted to a final concentration of 0.1% DMSO. The single base pair substitution of GCG (alanine) to TCG (serine) was introduced into the sensitive cDNA 14.1 by subcloning in a restriction fragment from the insensitive cDNA. The region containing the mutation was then amplified by PCR from total embryonic RNA from *Rdl*^{R-MD} flies and subsequently digested with *EcoRV* and *AclI*. This restriction fragment was then cloned into the sensitive cDNA 14.1 cut with the same enzymes to yield the Ala302 → Ser mutant construct. The presence of the mutation was confirmed by sequencing.

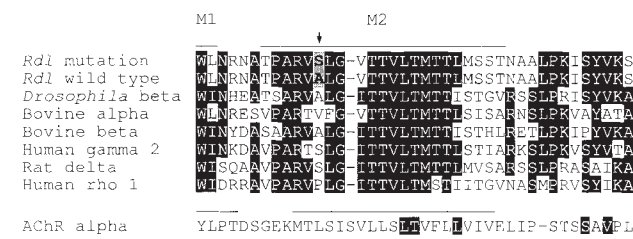


FIG. 1 Alignment of various GABA_A receptor subtype amino-acid sequences (single-letter code) and a nicotinic acetylcholine receptor (AChR) α -subunit²⁷ sequence showing the high degree of identity in M2. Amino acids found in five or more of the nine sequences are boxed in black. The amino-acid substitution between the wild-type and mutant *Rdl* sequence is shaded and indicated by an arrow. The position of the inner polar site¹⁵ within the AChR is indicated by an asterisk. GABA_A receptor sequences are *Drosophila* β ²⁸, bovine α 1, β 1²⁹, human γ 2³⁰, rat δ ³¹ and human ρ 1³².

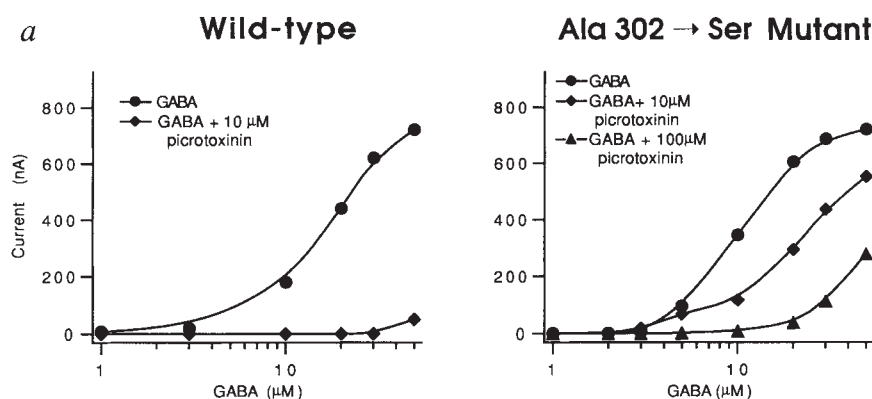
GABA receptors have been described previously in insects¹⁷ and similar insensitivity has also been observed in the recently identified GABA receptors in vertebrate retina, termed GABA_C receptors^{18,19}. In this context, it is interesting to note that homooligomeric expression of the ρ -subunit, which is highly expressed in the retina, has also been reported^{18,19}. The pharmacology of the ρ -homooligomeric receptor expressed in oocytes²⁰ appears to be very similar to that of the homooligomeric *Rdl* receptors. However, sequence comparison reveals that *Rdl* has no greater similarity to this than to any other known GABA subunit, and somewhat higher similarity to glycine receptors²¹. Therefore *Rdl* also appears to form a novel GABA receptor subtype.

Homooligomeric expression has also been described for other vertebrate¹⁶ and invertebrate²² GABA receptor subunits. In those studies, however, GABA-induced currents were small (<15 nA) and showed rapid desensitization. Coexpression with a second subunit was necessary before large GABA-evoked currents were observed. In contrast, in our study, currents of more than 1 μ A were frequently observed, and desensitization only occurred in response to high concentrations of GABA (>500 μ M). The fact that homooligomeric *Rdl* receptors form in oocytes does not necessarily mean that they are present *in vivo*. Further analysis of the composition of *Drosophila* GABA receptors awaits the characterization of other subunits. But the presence of homooligomeric receptors has recently been suggested as an explanation for the novel GABA responses observed in the vertebrate retina¹⁹.

To test the effect of the resistance-associated Ala302 $\rightarrow$ Ser mutation this substitution was introduced into the sensitive cDNA (Fig. 2). Channels expressed from this mutant construct were apparently identical to the wild-type channels in their

response to GABA and muscimol, but differed in their response to PTX and dieldrin. Pharmacological comparisons were made using 30 oocytes injected with each type of RNA. These oocytes were obtained from a number of different frogs and showed considerable variability in their response to GABA. But oocytes expressing RNA derived from the Ala302 $\rightarrow$ Ser mutant showed consistently lower sensitivity to PTX and dieldrin (Fig. 2b–e, Fig. 3). In oocytes injected with RNA derived from the wild-type 'sensitive' cDNA, application of 1 μ M PTX blocked more than 70% of the GABA response (Fig. 2b). This concentration of PTX had little effect in oocytes injected with RNA derived from the Ala302 $\rightarrow$ Ser mutant (Fig. 2d). It was necessary to increase the PTX concentration to 100 μ M to obtain a similar degree of block, indicating that the mutation induces roughly a 100-fold reduction in sensitivity. This is in agreement with the estimate of 100-fold reduction in sensitivity observed in cultured *Rdl*^{R-MD} neurons under whole-cell patch clamp (H.-G. Zang, R.H.f.C. and M. Jackson, unpublished data).

The sensitivity to the cyclodiene insecticide dieldrin was also reduced. Thus, 10 μ M dieldrin caused >75% block of the GABA response in oocytes expressing the wild-type 'sensitive' cDNA (Fig. 2c), but had little effect on those expressing the insensitive mutant cDNA (Fig. 2e). Further increase of the dieldrin concentration was not possible because of its limited solubility, but a comparison of the effects of lower doses suggests that the mutation confers 50–100-fold loss of sensitivity to dieldrin. Insensitivity to this concentration of dieldrin is consistent with previous suction electrode recordings taken from the larval central nerve cord of *Rdl*^{R-MD} (ref. 23). Although resistance levels for both PTX and dieldrin may be lower than those observed in bioassays with adult flies, this can be explained by the difficulty of killing more than 50% of homozygous resistant



b

Reduction in GABA response (%) (mean (s.d.))

GABA EC ₅₀	1 μ M PTX		10 μ M PTX		50 μ M PTX		10 μ M dieldrin	
	Wt	Mutant	Wt	Mutant	Wt	Mutant	Wt	Mutant
10 μ M	73 (15) <i>n</i> =3	8 (10) <i>n</i> =6	96 (4) <i>n</i> =3	61 (14) <i>n</i> =7	96 (5) <i>n</i> =4	75 (28) <i>n</i> =3	78 (4) <i>n</i> =2	23 (21) <i>n</i> =3
50 μ M	78 (12) <i>n</i> =9	8 (7) <i>n</i> =7	90 (13) <i>n</i> =6	48 (6) <i>n</i> =4	97 (5) <i>n</i> =5	56 (6) <i>n</i> =5	80 (6) <i>n</i> =4	12 (11) <i>n</i> =6

FIG. 3 *a*, Examples of response of individual oocytes to varying concentrations of GABA and PTX. GABA was perfused until maximum response was obtained, then washed out for 5 min before application of the next higher concentration. Subsequently, PTX or dieldrin was perfused for 10 min, then mixtures of GABA and PTX applied. At the end of the experiment, the preparation was washed with saline for 20 min and the response to a high dose of GABA measured. In all experiments, the response to GABA recovered

by at least 50% during washing. Complete dose-response curves were obtained for three oocytes of each type, and were similar in shape despite differences in the threshold concentration for the GABA effect. *b*, Additional oocytes were tested by finding the rough EC₅₀ for GABA, then applying one or more concentrations of PTX or dieldrin (some oocytes were treated with more than one concentration). 25 Of 30 oocytes had a GABA EC₅₀ between 10 and 50 μ M, and these data are summarized here.

flies in contact bioassays at any dose²⁴.

The finding that mutation of a single amino acid can confer significant insensitivity to this receptor has implications both for structure-function studies of ion channels and for the molecular evolution of pesticide resistance. The location of the substitution at the cytoplasmic (inner) end of M2 in *Rdl* is analogous (Fig. 1) to the location of the inner polar site of the nicotinic acetylcholine receptor (AChR). Experiments involving the substitution of serines to alanines within the inner polar site of the AChR result in alteration of interactions both with permeant cations and with an open-channel blocker and have led to the conclusion that this part of the M2 region is within the ion channel pore¹⁵. Residues within M2 of the closely related glycine receptor have also been identified as major determinants of PTX resistance by site-directed mutagenesis²⁵, but no specific residue was singled out as having a major effect. The alanine to serine substitution in *Rdl*, although not introducing any change in the net charge, increases the polarity of M2 by the addition of a hydroxyl group and may result in steric or electrostatic hindrance of PTX with its binding site in the channel pore. The amino-acid sequence within M2 is highly conserved and either an alanine or a serine is found at the equivalent position in some other GABA receptor subunits (Fig. 1). Therefore it will also be interesting to examine the role of this amino-acid position in determining PTX sensitivity in other GABA receptors where coassembly of a number of different subunits is required for functional expression.

The finding that cyclodiene resistance, widespread in insects¹³ and also found in vertebrates⁹, can be conferred by a single amino-acid substitution has important implications, for both the monitoring and evolution of resistance. Knowledge of the single substitution responsible for resistance greatly aids the genotyping of individual insects in resistance monitoring techniques. To this end, we have developed a technique based on polymerase chain reaction (PCR) amplification of genomic DNA from the resistance gene by which individual flies can be identified as homozygous susceptible (*Rdl*^S/*Rdl*^S), homozygous resistant (*Rdl*^R/*Rdl*^R) or heterozygous (*Rdl*^R/*Rdl*^S) by the presence or absence of a diagnostic restriction enzyme site¹².

In the context of resistance evolution there has been considerable debate over the importance of monogenic versus polygenic mechanisms of resistance²⁶. The finding that cyclodiene resistance is conferred not only by a single gene but, in *Drosophila*, globally by a similar single amino-acid substitution within that gene¹², strongly supports the significance of single major genes in the evolution of resistance. Further, in view of the highly conserved region of the protein in which the mutation is found, the possibility is raised that a similar amino-acid substitution may be found in other cyclodiene-resistant insects. We will investigate this hypothesis by sequence analysis of a range of *Rdl* homologues in cyclodiene-sensitive and -insensitive insects. Finally, it is interesting to note that a mutation in such a functional part of a protein can be maintained in field populations in the apparent absence of selection pressure. We will thus perform a detailed pharmacological analysis of the mutant receptor to establish if normal channel function is impaired. □

12. ffrench-Constant, R. H., Steichen, J., Rocheleau, T. A., Aronson, K. & Roush, R. T. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
13. Georgioli, G. P. in *Pesticide Resistance: Strategies and Tactics for Management* 14–43 (National Academy Press, Washington DC, 1986).
14. Imoto, K. *et al. Nature* **335**, 645–648 (1988).
15. Leonard, R. J., Labarca, C. G., Charnet, P., Davidson, N. & Lester, H. A. *Science* **242**, 1578–1581 (1988).
16. Blair, L. A., Levitan, E. S., Marshall, J., Dionne, V. E. & Barnard, E. A. *Science* **242**, 577–579 (1988).
17. Lummis, S. C. *Comp. Biochem. Physiol.* **95**, 1–8 (1990).
18. Feigenspan, A., Wassle, H. & Bormann, J. *Nature* **361**, 159–161 (1993).
19. Qian, H. & Dowling, J. E. *Nature* **361**, 162–164 (1993).
20. Shimada, S., Cutting, G. & Uhl, G. R. *Molec. Pharmacol.* **41**, 683–687 (1992).
21. ffrench-Constant, R. H. & Rocheleau, T. *J. Neurochem.* **59**, 1562–1565 (1992).
22. Harvey, R. J. *et al. EMBO J.* **10**, 3239–3245 (1991).
23. Bloomquist, J. R., ffrench-Constant, R. H. & Roush, R. T. *Pestic. Sci.* **32**, 463–469 (1991).
24. ffrench-Constant, R. H. & Roush, R. T. *Genet. Res. Camb.* **57**, 17–21 (1991).
25. Pribilla, I., Takagi, T., Langosch, D., Bormann, J. & Betz, H. *EMBO J.* **11**, 4305–4311 (1992).
26. Roush, R. T. & McKenzie, J. A. *Rev. Ent.* **32**, 361–380 (1987).
27. Noda, M. *et al. Nature* **299**, 793–797 (1982).
28. Satelle, D. B. *et al. in Transmitter Amino Acid Receptors, Structures, Transduction and Models for Drug Development* (eds Barnard, E. A. & Costa, E.) 273–291 (Thieme Medical, New York, 1992).
29. Schofield, P. R. *et al. Nature* **328**, 221–227 (1987).
30. Pritchett, D. B. *et al. Nature* **338**, 582–585 (1989).
31. Shivers, B. D. *et al. Neuron* **3**, 327–337 (1989).
32. Cutting, G. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 2673–2677 (1991).

ACKNOWLEDGEMENTS. We thank M. A. Wolff for technical assistance, and M. Jackson, M. Tomalski, J. Thackeray and B. Ganetzky for comments on the manuscript. This work was supported by grants from the NIH and Hatch to R. ff.-C. and by Rhone Poulenc Ag Company, and was initiated in the laboratory of R. T. Roush.

Endogenous dynorphins inhibit excitatory neurotransmission and block LTP induction in the hippocampus

John J. Wagner*, Gregory W. Terman† & Charles Chavkin‡

Departments of Pharmacology and †Anesthesiology, SJ-30, University of Washington, Seattle, Washington 98195, USA

ALTHOUGH anatomical and neurochemical studies suggest that endogenous opioids act as neurotransmitters^{1–7}, their roles in normal and pathophysiological regulation of synaptic transmission are not defined. Here we examine the actions of prodynorphin-derived opioid peptides in the guinea-pig hippocampus and show that physiological stimulation of the dynorphin-containing dentate granule cells can release endogenous dynorphins, which then activate κ_1 opioid receptors present in the molecular layer of the dentate gyrus. Activation of κ_1 receptors by either pharmacologically applied agonist or endogenously released peptide reduces excitatory transmission in the dentate gyrus, as shown by a reduction in the excitatory postsynaptic currents evoked by stimulation of the perforant path, a principal excitatory afferent. In addition, released dynorphin peptides were found to block the induction of long-term potentiation (LTP) at the granule cell-perforant path synapse. The results indicate that endogenous dynorphins function in this hippocampal circuit as retrograde, inhibitory neurotransmitters.

Whole cell recordings were made from granule cells in the guinea-pig hippocampal slice (Fig. 1) and excitatory postsynaptic currents (e.p.s.cs) were evoked by afferent stimulation either in the molecular layer (to activate perforant path fibres from the entorhinal cortex) or in the hilus (to activate commissural/associational afferents). Opioid receptor activation by the κ_1 selective agonist U69,593 (refs 8, 9) at 500 nM significantly reduced (by $41 \pm 3\%$; $n = 8$) the amplitude of the CNQX-sensi-

Received 3 December 1992; accepted 15 March 1993.

1. Kuffler, S. W. & Edwards, C. J. *Neurophysiol.* **21**, 589–601 (1965).
2. Usherwood, P. N. R. & Grundfest, H. *J. Neurophysiol.* **28**, 497–518 (1965).
3. Olsen, R. W. & Tobin, A. J. *FASEB J.* **4**, 1469–1480 (1990).
4. Eldefrawi, A. T. & Eldefrawi, M. E. *FASEB J.* **1**, 262–271 (1987).
5. Meldrum, B. S. *Br. J. clin. Pharmacol.* **27**, 3S–11S (1989).
6. Wagstaff, J. *et al. Am. J. hum. Genet.* **49**, 330–337 (1991).
7. Matsumura, F. & Ghisuddin, S. M. *J. environ. Sci. Hlth* **B18**, 1–14 (1983).
8. Kadous, A. A., Ghisuddin, S. M., Matsumura, F., Scott, J. G. & Tanaka, K. *Pestic. Biochem. Physiol.* **19**, 157–166 (1983).
9. Yarbrough, J. D., Roush, R. T., Bonner, J. C. & Wise, D. A. *Experientia* **42**, 851–853 (1986).
10. ffrench-Constant, R. H., Roush, R. T., Mortlock, D. & Dively, G. P. *J. Ecol. Ent.* **83**, 1733–1737 (1990).
11. ffrench-Constant, R. H., Mortlock, D. P., Shaffer, C. D., MacIntyre, R. J. & Roush, R. T. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7209–7213 (1991).

* Present address: Department of Physiology, University of Maryland School of Medicine, Baltimore, MD 21201, USA.

‡ To whom correspondence should be addressed.

tive e.p.s.cs evoked by perforant path stimulation without affecting granule cell input conductance (Fig. 1*b*). This effect was reversed by the κ_1 selective antagonist^{10,11}, norbinaltorphimine at 100 nM (Fig. 1*b*). In contrast, the CNQX-sensitive component of the e.p.s.cs evoked by hilar stimulation was not significantly affected by κ_1 receptor activation ($n=4$) (Fig. 1*b*). Thus, κ_1 receptors appear to be selectively expressed on perforant path terminals and to inhibit glutamate release rather than directly affecting the postsynaptic cell or its response to glutamate.

To determine whether endogenous opioids also modulate the release of glutamate from perforant path afferents, we stimulated granule cells using a paradigm previously shown to release dynorphins by antidromic activation of granule cell axons in the hilus of the dentate gyrus⁶. Perforant path-evoked e.p.s.cs were monitored before and after dynorphin release, and e.p.s.c. amplitudes were found to be significantly reduced (by $21 \pm 2\%$, $n=15$) following hilar stimulation. The onset of e.p.s.c. inhibition was evident in the first minute after antidromic stimulation and was maximal by 1.7 ± 0.1 min post-stimulation (range, 1.0–2.3 min; $n=15$). For the representative granule cell recording shown in Fig. 2*a*, e.p.s.c. amplitude was reduced by $\sim 24\%$ following a hilar stimulus train, and a second stimulus train to the hilus evoked a similar (27%) depression in e.p.s.c. response. The duration of the inhibition was more variable (Fig. 2).

The reduction in e.p.s.c. amplitude caused by hilar stimulation was blocked by $1 \mu\text{M}$ naloxone ($0 \pm 5\%$ change; $n=5$) at 2 min after hilar high-frequency stimulation (HHFS). In the representative cell shown, hilar stimulation reduced perforant path e.p.s.cs by 26%, whereas in the presence of naloxone, perforant

TABLE 1 Effects of NBNI and dynorphin antisera on perforant path-evoked granule cell population response amplitude after high-frequency stimulation of the hilus and/or perforant path

	Per cent change from prestimulation spike amplitude		
	1 min post-HHFS	Perforant path LTP	HHFS/perforant path LTP
Control buffer	-31.9 ± 3.5 (29)	52.0 ± 6.7 (13)	-5.1 ± 3.3 (9)
100 nM NBNI	-11.4 ± 4.4 (9)*	40.6 ± 11.6 (6)	34.7 ± 5.3 (8)*
Anti-dynorphin	-6.5 ± 3.7 (6)*	ND	44.1 ± 11.8 (5)*

Data are expressed as per cent change from the mean population spike amplitude measured during the five minutes before high-frequency stimulation. In control buffer, high-frequency stimulation of the hilus reduced the amplitude of the population spike by 32% (1 min post-HHFS column). This effect was significantly attenuated by pretreatment with 100 nM norbinaltorphimine (NBNI) or dynorphin antisera (anti-dynorphin) 15 min before hilar high-frequency stimulation. Perforant path LTP was induced by high-frequency stimulation of the perforant path, resulting in a 52% increase in response to low-frequency perforant path stimulation 30 min post-induction (mean amplitude 25–30 min after high-frequency stimulation). NBNI had no effect under these conditions. Administering the hilar high-frequency stimulation immediately before the perforant path high-frequency stimulation (HHFS/perforant path LTP column) eliminated any potentiation measured 30 min after stimulation in control buffer. Pretreatment with either NBNI or dynorphin antisera resulted in significant LTP following perforant path high-frequency stimulation in spite of preceding hilar high-frequency stimulation. Control buffer data includes slices treated with normal rabbit sera (1:125) 15 min before high-frequency stimulation. Responses of this treatment group showed no significant differences from slices in normal buffer. Population spike amplitudes were chosen as the most robust measure of the changes observed; field e.p.s.p. slopes were also measured and showed the same κ -mediated inhibition of synaptic transmission and LTP. Data are means $\pm$ s.e.m. from the number of independent experiments given in parentheses. Asterisks indicate statistical difference from similarly stimulated control buffer group ($P < 0.05$). ND, not determined.

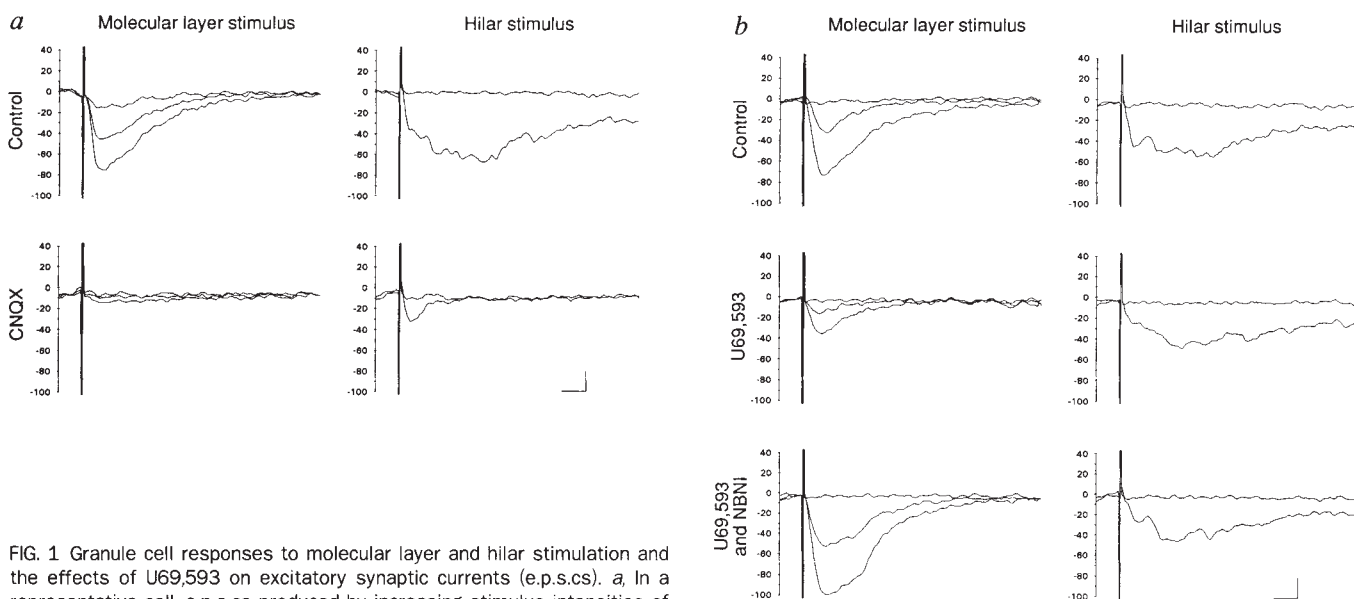


FIG. 1 Granule cell responses to molecular layer and hilar stimulation and the effects of U69,593 on excitatory synaptic currents (e.p.s.cs). *a*, In a representative cell, e.p.s.cs produced by increasing stimulus intensities of perforant path (20, 25, 30 μA) or hilar (100, 150 μA) stimulation in the presence of bicuculline (10 μM) were both largely blocked by 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; 5 μM). U69,593 affected neither the membrane conductance of granule cells at membrane potentials from -120 to -60 mV (data not shown), nor the holding current required to clamp the cell at -70 mV. *b*, In another representative granule cell in the presence of bicuculline, the κ opioid agonist U69,593 (500 nM) inhibited CNQX-sensitive e.p.s.cs from from perforant path. This inhibition was completely reversed by the κ antagonist NBNI (100 nM; $n=4$). The slight reduction in e.p.s.c. amplitude in the presence of U69,593 seen after hilar stimulation was neither reversed by NBNI nor statistically significant in 4 replicates ($P > 0.05$). Individual traces shown in *a* and *b* are the averages of two sequential sweeps; the experiment was done 3 times with similar results. Sweep length is 110 ms; vertical scales are current amplitudes (pA); calibration is 10 ms (horizontal), 40 pA (vertical) for all traces.

METHODS. Guinea-pig hippocampal slices (500 μm) were prepared as described⁷ and were perfused with Krebs bicarbonate buffer containing (mM): NaCl (125), KCl (3), CaCl_2 (2), MgCl_2 (1), NaH_2PO_4 (1.25), NaHCO_3 (26), and glucose (10), saturated with 95% O_2 , 5% CO_2 , pH 7.4. Patch pipettes

(3–5 M Ω) contained (mM): CsCl (120), CaCl_2 (1), MgCl_2 (2), ATP (5), tetraethylammonium-Cl (20), EGTA (10), HEPES (10), adjusted to pH 7.2 with CsOH. Whole-cell voltage clamp recordings were obtained using an Axopatch 200 amplifier and analysed using FastLab software. Uncompensated series resistance was checked throughout the recording period, and the data excluded from analysis if the drift exceeded 20%. To study isolated e.p.s.cs, bicuculline (10 μM) was added to the bath to block GABA_A receptors in all studies reported here and CsCl was included in the intracellular recording pipettes to block GABA_B receptor-gated potassium currents. Spontaneous activity was virtually eliminated after GABA_A receptors were blocked by bicuculline (45 of 45 cells). For statistical analysis we used analysis of variance with or without repeated measures as appropriate. Tukey's tests were used for *post-hoc* comparisons; $P < 0.05$ was considered to be significant.

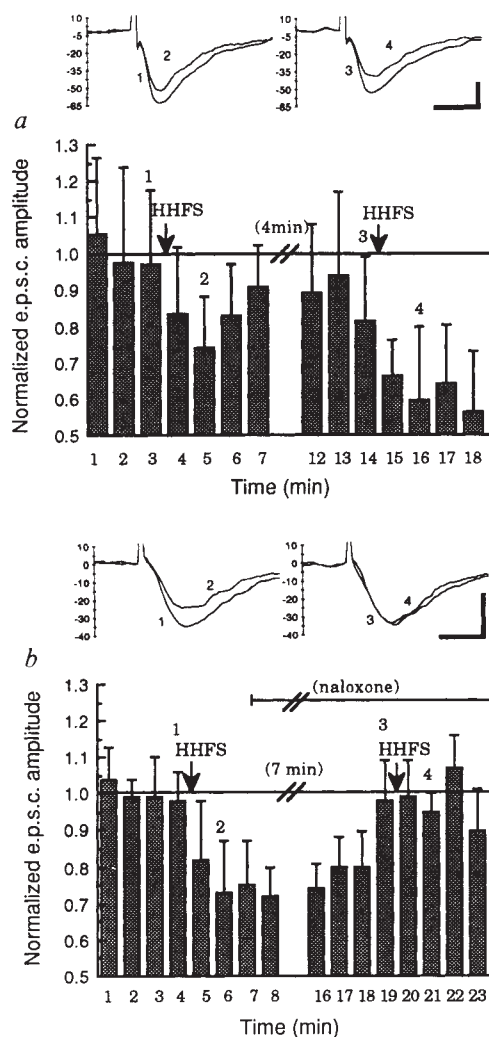


FIG. 2 Hilar high-frequency stimulation (HHFS) induces inhibition of perforant path-evoked excitatory postsynaptic potentials (PP e.p.s.cs) recorded in the granule cell. **a**, HHFS causes a 24% reduction in the amplitude of the granule cell e.p.s.cs. Following recovery (>90% of control) from the initial HHFS, a second HHFS was again able to inhibit PP e.p.s.c. amplitude (compare sweeps 3 and 4). **b**, HHFS causes a 26% reduction in the PP e.p.s.c. amplitude (compare sweeps 1 and 2). Twelve minutes after addition of 1 μ M naloxone to the superfusion buffer, the PP e.p.s.c. amplitude returned to a pre-HHFS level and another HHFS reduced PP e.p.s.c. amplitude by only 4% (compare sweeps 3 and 4). Vertical scales for figure insets are current amplitudes (pA). Stimulus artefacts are truncated. Group data: HHFS significantly decreased e.p.s.c. amplitude at two ($-15 \pm 2\%$), three ($-13 \pm 3\%$) and four ($-16 \pm 3\%$) min post-stimulus in normal Krebs bicarbonate buffer ($n=15$); it had no effect at any time point when naloxone (1 μ M) was added to the perfusate at least 10 min before HHFS ($n=5$).

METHODS. The opioid-mediated effects of HHFS were monitored by measuring granule cell e.p.s.c. amplitudes evoked by a perforant path test pulse in the presence or absence of naloxone. PP e.p.s.cs were elicited at 0.1 Hz and 6 sweeps averaged into 1 min bins (bars are means $\pm$ s.e.m. of the 6 sweeps). For each cell tested, the mean of 3 or 4 min of pre-HHFS e.p.s.c. amplitudes was determined, and the per cent of that control value calculated. Hilar stimulation at high frequency (50 Hz, 1 s train of 0.3 ms, 150- μ A pulses) was applied using a concentric bipolar electrode (Kopf, SNE 100). Calibration is 10 ms (horizontal), 25 pA (vertical) for all traces.

path e.p.s.c. amplitude was reduced only 4% following hilar stimulation (Fig. 2b). This antagonism of the inhibitory effects of hilar stimulation indicates that, under these conditions, endogenously released opioids, like exogenously applied opioids, can regulate excitatory neurotransmission in the dentate gyrus.

The temporal characteristics of the inhibitory effects of endogenous opioids were further defined using population spike amplitude measurements because this extracellular response is more stable. Antidromic granule cell activation by high-frequency hilar stimulation also significantly inhibited the population response evoked by glutamate release from the perforant path (Fig. 3a). The onset was within 2 min and the duration was 4.1 ± 0.4 min (range, 2 to 8 min; $n=17$) following hilar stimulation. The κ_1 antagonist norbinaltorphimine at 100 nM also significantly blocked the reduction in population response amplitudes caused by hilar stimulation (Fig. 3b).

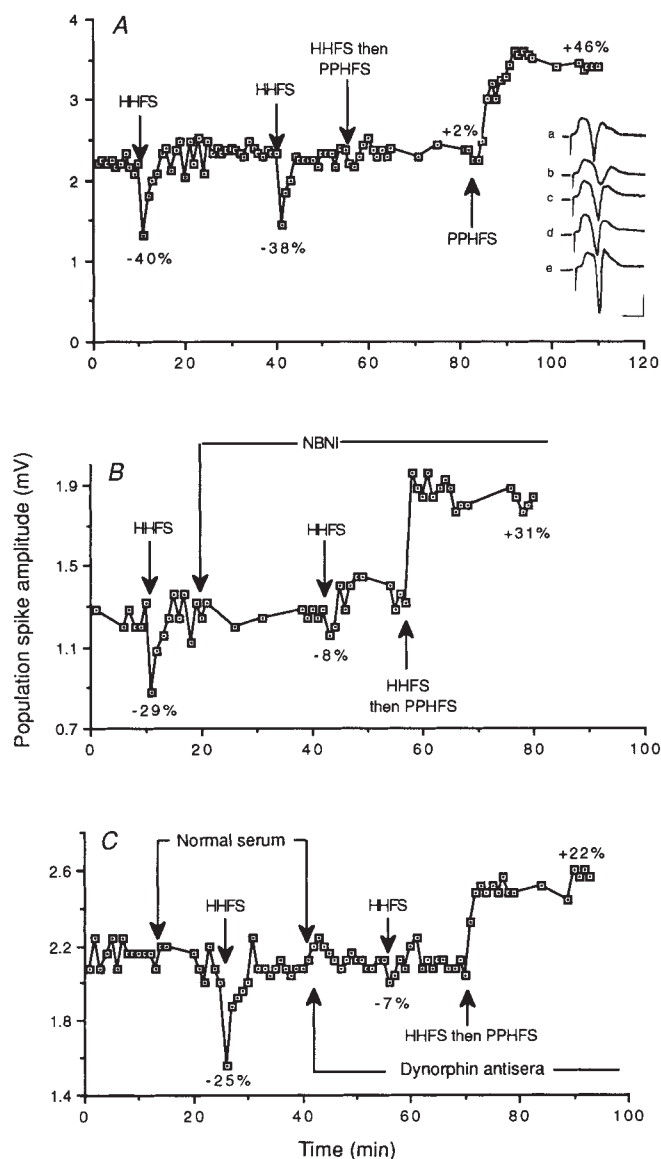
To identify the opioid peptide responsible for the inhibition of granule cell excitability, the effects of polyclonal antisera⁶ raised against the prodynorphin-derived opioid peptides were tested (antisera were a gift from C. Evans). A mixture of antisera against dynorphin A(1-8), dynorphin B, and α -neoendorphin was found significantly to antagonize the hilar stimulation-evoked inhibitory effect when added to the buffer perfusing the hippocampal slices (Fig. 3c). Population spike amplitude was not directly affected in the absence of hilar stimulation by norbinaltorphimine, dynorphin antisera, or control rabbit antisera (Fig. 3). The ability of dynorphin selective antisera to block hilar stimulation-induced inhibition of synaptic transmission supports the conclusion that endogenous dynorphins are released to suppress the response evoked by perforant path stimulation.

The inhibitory effects of endogenous κ opioids on excitatory input to the granule cell indicated that opioid peptides might also modulate long-term sequelae from such afferent input. High-frequency perforant path stimulation consistently produced long-term potentiation; but if the perforant path stimulation train was immediately preceded by hilar high-frequency stimulation, LTP production was blocked (Fig. 3a). This blockade was attenuated by either norbinaltorphimine (100 nM; Fig. 3b) or dynorphin antisera (Fig. 3c), indicating that endogenous dynorphins released by hilar stimulation can act through κ_1 receptors to inhibit LTP at the perforant path-granule cell synapse. Data from a series of replicates are summarized in Table 1.

Our results demonstrate an inhibitory effect of endogenously released dynorphins on excitatory neurotransmission and synaptic plasticity in the dentate gyrus of the guinea-pig hippocampus. Functionally, the dynorphins released by high-frequency stimulation appear to act as retrograde transmitters, providing negative feedback to the presynaptic terminal. The sites of dynorphin release were not defined by these experiments, but are likely to be either from granule cell dendrites or recurrent collaterals¹⁻⁴. Following stimulated release, the effects of endogenous dynorphins were slow and prolonged, reaching a peak within 2 min and lasting for several minutes. These temporal characteristics are similar to those of other neuropeptides acting in the peripheral nervous system^{12,13}.

The finding that dynorphin can block LTP production suggests that this peptide powerfully affects synaptic plasticity and, presumably, learning and memory, for which the phenomenon of LTP has been proposed to be a cellular substrate¹⁴⁻¹⁶. Indeed, exogenous opioids inhibit learning and memory^{17,18}. Elevated hippocampal dynorphin levels correlate with impaired spatial learning in aged rats¹⁹, an impairment attenuated by naloxone²⁰. A naloxone-sensitive learning deficit can also be produced by mossy fibre stimulation²¹, similar to that reported here. Nonetheless, the role of opioid peptides in modulating learning and memory is likely to be complicated as μ and δ opioids (which suppress GABA, not glutamate, release) facilitate LTP induction

FIG. 3 Both norbinaltorphimine and dynorphin antisera block the inhibitory effects of hilar high-frequency stimulation on granule cell population spike amplitude and long-term potentiation produced by perforant path high-frequency stimulation on granule cell population spike amplitude and long-term potentiation produced by perforant path high-frequency stimulation (PPHFS). A, Representative experiment showing the effects of HHFS, HHFS followed immediately by PPHFS, and PPHFS on granule cell population spike responses. HHFS (6 1-s, 50 Hz trains of 0.3-ms 300 μ A pulses given at a rate of 1 every 10 s) decreased spike amplitude by 40% and then 38%, showing nearly identical inhibitory effects of repeated HHFS. Whereas HHFS followed immediately by PPHFS resulted in a minimal change (+2%) in spike amplitude 30 min after stimulation, PPHFS by itself produced a 46% increase 30 min after stimulation (that is, LTP). Inset shows representative traces from a replicate experiment at different times during the protocol: a, control; b, 1 min after HHFS; c, 20 min after HHFS; d, 1 min after HHFS immediately followed by PPHFS; and e, 30 min after PPHFS-induced LTP. Responses were evoked at 55 μ A (the $S_{1/2}$, or half-maximal spike amplitude, for this preparation); calibration bars for the inset are 5 ms (horizontal), 1 mV (vertical). B, Effect of 100 nM NBNI on the HHFS-induced attenuation of the population spike and PPHFS-evoked LTP. HHFS produced a 29% decrease in spike amplitude initially, but only an 8% decrease after NBNI was added to the perfusate. Unlike the results shown in A, HHFS given immediately before PPHFS in the presence of NBNI did not block LTP (+31%). C, Effect of dynorphin antisera on HHFS-induced reduction of the population spike and LTP following PPHFS. With normal rabbit serum (1:125) in the perfusate, HHFS decreased spike amplitude 25% (similar in magnitude to A and B). After dynorphin antisera was added to the perfusate, HHFS reduced spike amplitude by only 7% and, unlike the results shown in A, a 22% potentiation was seen with HHFS immediately followed by PPHFS. METHODS. Extracellular recordings were made under the conditions described in Fig. 1 legend, except that the concentrations of $CaCl_2$ and $MgCl_2$ in the extracellular buffer were each increased to 4 mM to inhibit hyperexcitability in the presence of the 10 μ M of bicuculline. Recording pipettes were filled with 3 M NaCl. The perforant path stimulation-induced dentate granule cell population response amplitude was measured from peak to peak using a digitizing oscilloscope. A stimulus intensity was chosen that evoked a half-maximal granule cell population spike amplitude ($S_{1/2}$). The LTP induction paradigm consisted of a total of 9 pulses (0.3 ms, 300 μ A) to the perforant path given as three trains (each with three stimuli) applied at 10-s intervals with 10 ms between stimuli. LTP was assessed at 25–30 min post-PPHFS. Pooled dynorphin antisera included rabbit polyclonal antisera raised against dynorphin A_{1-8} , dynorphin B and α -neoendorphin. The preparation and characterization of this antisera has been described^{6,30}. All three antisera were added to the perfusate at the following dilutions: 1:300, dynorphin A_{1-8} ; 1:250, dynorphin B; 1:1,000, α -neoendorphin (for a final titre of 1:120 serum in the medium).



in the dentate gyrus^{22,23}.

The regulation of dentate gyrus excitability by dynorphin may also be important in seizure disorders. κ agonists inhibit seizure activity in several animal models^{24,25}; mossy fibre stimulation can cause a naloxone-reversible elevation in seizure threshold²⁶; and seizures have been correlated in humans²⁷ and in animal models^{28,29} with increases in granule cell dynorphin expression and mossy fibre sprouting. Thus dynorphin, by providing feedback control of afferent input to the hippocampus, may normally act to modulate hippocampal activity during excitation involved in learning and during pathological excitation resulting in seizures or cell death.

Note added in proof: A report on a similar topic was submitted and published³¹ since submission of this letter. □

Received 29 December 1992; accepted 29 April 1993.

- Gall, C., Brecha, N., Karten, H. J. & Chang, K. *J. comp. Neurol.* **198**, 335–350 (1981).
- McGinty, J. F., Henriksen, S. J., Goldstein, A., Terenius, L. & Bloom, F. E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 589–593 (1983).
- Chavkin, C., Shoemaker, W. J., McGinty, J. F., Bayon, A. & Bloom, F. E. *J. Neurosci.* **5**, 808–816 (1985).
- McLean, S., Rothman, R., Jacobson, A., Rice, K. & Herkenham, M. *J. comp. Neurol.* **255**, 497–510 (1987).
- Chavkin, C., Bakhtit, C., Weber, E. & Bloom, F. E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7669–7673 (1983).
- Wagner, J. J., Evans, C. & Chavkin, C. *J. Neurochem.* **57**, 333–343 (1991).

- Wagner, J. J., Caudle, R. M. & Chavkin, C. *J. Neurosci.* **12**, 132–141 (1992).
- Lahti, R., Mickelson, M., McCall, J. & von Voigtlander, P. *Eur. J. Pharmac.* **109**, 281 (1985).
- Zukin, R., Eghball, M., Olive, D., Unterwald, E. & Tempel, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4061–4065 (1988).
- Portoghesi, P. S., Lipkowski, A. W. & Takemori, A. E. *Life Sci.* **40**, 1287–1292 (1987).
- Nock, B., Giordano, L., Cicero, T. & O'Connor, L. *J. Pharmac. exp. Ther.* **254**, 412–419 (1990).
- Jan, L. Y. & Jan, Y. N. *J. Physiol., London* **327**, 219–246 (1982).
- Lundberg, J. M., Rudehill, A., Sollevi, A., Theodorsson-Norheim, E. & Hamberger, B. *Neurosci. Lett.* **63**, 96–100 (1986).
- Bliss, T. V. P. & Lomo, T. *J. Physiol.* **232**, 331 (1973).
- Lynch, G. & Baudry, M. *Science* **224**, 1057–1063 (1984).
- Madison, D. V., Malenka, R. C. & Nicoll, R. A. *Rev. Neurosci.* **14**, 379–397 (1991).
- Martinez, J. L., Janak, P. H., Weinberger, S. G., Schulties, G. & Derrick, R. E. in *NIDA Res. Monogr.* **97** (ed. Erinoff, L.) 48–78 (US Govt Printing Office, Washington DC, 1990).
- Decker, M. W. & McGaugh, J. L. *Synapse* **7**, 151–168 (1991).
- Jiang, H. K., Owyang, V., Hang, J. S. & Gallagher, M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2948–2951 (1989).
- Gallagher, M., Bostock, E. & King, R. A. *Behav. Neural Biol.* **44**, 374–385 (1985).
- Collier, T. J. & Routtenberg, A. *Brain Res.* **310**, 384–387 (1984).
- Bramham, C., Milgram, N. & Srebo, B. *Brain Res.* **567**, 42–50 (1991).
- Xie, C. W. & Lewis, D. V. *J. Pharmac. exp. Ther.* **256**, 289–296 (1991).
- Tortella, F. C. *Trends pharmac. Sci.* **9**, 366–372 (1988).
- Singh, L., Vass, C. A., Hunter, J. C., Woodruff, G. N. & Hughes, J. *Eur. J. Pharmac.* **191**, 477–480 (1990).
- Jones, L. S. *Brain Res.* **564**, 336–340 (1991).
- Houser, C. R., Miyashiro, J. E., Walsh, G. O., Rich, J. R. & Delgado-Escueta, A. V. *J. Neurosci.* **10**, 267–282 (1990).
- Sloviter, R. S. *Neurosci. Lett.* **137**, 91–96 (1992).
- Cronin, J., Obenaus, A., Houser, C. R. & Dudek, F. E. *Brain Res.* **573**, 305–310 (1992).
- Evans, C., Erdelyi, E., Hunter, J. & Barchas, J. *J. Neurosci.* **5**, 3423–3427 (1985).
- Weisskopf, M. G., Zalutsky, R. A. & Nicoll, R. A. *Nature* **362**, 423–427 (1993).

ACKNOWLEDGEMENTS. We thank C. Evans for dynorphin antisera. The work was supported by the National Institute on Drug Abuse. G.W.T. was supported by a NRSA grant from NIGMS.

Functional expression of a rapidly inactivating neuronal calcium channel

Patrick T. Ellinor*, Ji-Fang Zhang*, Andrew D. Randall*, Mei Zhou†‡, Thomas L. Schwarz*, Richard W. Tsien*§ & William A. Horne†‡

* Department of Molecular and Cellular Physiology, Stanford University Medical Center, Stanford, California 94305, USA

† Department of Pharmacology, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

‡ Present address: Neurex Corporation, 3760 Haven Avenue, Menlo Park, California 94025, USA.

§ To whom correspondence should be addressed.

DIVERSE types of calcium channels in vertebrate neurons are important in linking electrical activity to transmitter release, gene expression and modulation of membrane excitability¹. Four classes of Ca^{2+} channels (T, N, L and P-type) have been distinguished²⁻⁶ on the basis of their electrophysiological and pharmacological properties. Most of the recently cloned Ca^{2+} channels⁷⁻¹⁶ fit within this functional classification. But one major branch of the Ca^{2+} channel gene family, including BII (ref. 15) and doe-1 (ref. 16), has not been functionally characterized. We report here the expression of doe-1 and show that it is a high-voltage-activated (HVA) Ca^{2+} channel that inactivates more rapidly than previously expressed calcium channels. Unlike L-type or P-type channels, doe-1 is not blocked by dihydropyridine antagonists or the peptide toxin ω -Aga-IVA, respectively. In contrast to a previously cloned N-type channel¹⁴, doe-1 block by ω -CTx-GVIA requires micromolar toxin and is readily reversible. Unlike most HVA channels, doe-1 also shows unusual sensitivity to block by Ni^{2+} . Thus, doe-1 is an HVA Ca^{2+} channel with novel functional properties. We

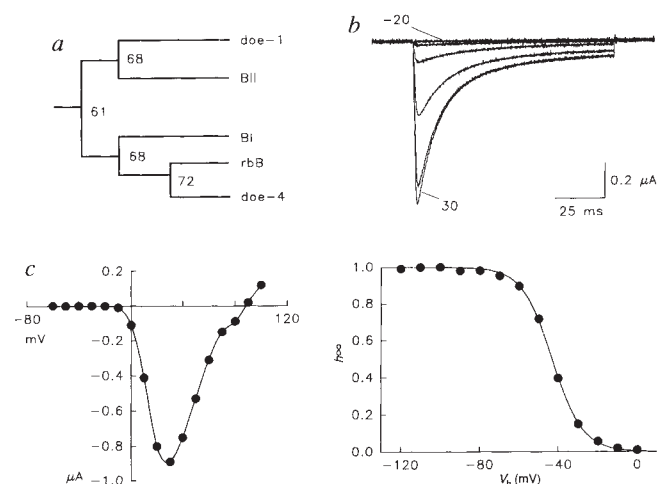
have identified a Ca^{2+} channel current in rat cerebellar granule neurons that resembles doe-1 in many kinetic and pharmacological features.

Voltage-dependent Ca^{2+} channels exist as a complex of subunits, including α_1 , α_2 , β , γ and δ , but most of the key functional properties of Ca^{2+} channels are attributed to α_1 subunit^{17,18}. Molecular cloning has identified two structural subfamilies of α_1 subunits^{2,6}. One subfamily consists of L-type channels, largely derived from various muscle tissues and characterized by sensitivity to dihydropyridine (DHP) compounds⁷⁻¹⁰. Another subfamily (Fig. 1a) consists of channels isolated from vertebrate brain¹¹⁻¹⁶. The clones for only one branch of this subfamily have been expressed; they (BI and rbB) lack sensitivity to DHPs^{11,12}. The second branch (Fig. 1a, top) contains two channels: BII, a Ca^{2+} channel cloned from rabbit brain¹⁵, and doe-1, a channel we recently isolated from the forebrain of *Discopyge ommata*¹⁶. Synaptic terminals of this marine ray have provided a rich source of neuronal Ca^{2+} channels^{19,20} and synaptic vesicle proteins^{21,22}.

Microinjection into *Xenopus* oocytes of complementary RNAs encoding doe-1, α_2 and β -subunits produced a robust Ba^{2+} current with rapid activation and inactivation kinetics. Figure 1b illustrates Ba^{2+} currents (40 mM external Ba^{2+}) elicited by steps from a holding potential (V_h) of -80 mV to various test potentials (V_t). The peak current is maximal at $+30$ mV (Fig. 1c), indicating that doe-1 encodes an HVA channel. The steady-state inactivation curve, obtained by varying V_h with $V_t = +20$ mV, is well-fitted by a Boltzmann relationship with a mid-point of -46.0 ± 2.5 mV (mean $\pm$ s.d., $n = 4$, Fig. 1d).

The doe-1 current activates and inactivates more rapidly than either BI or the cardiac-derived L-type channel. Figure 2a compares the currents obtained from oocytes in which either doe-1, cardiac L-type⁸, or BI¹¹ α_1 subunits were coexpressed with the same auxiliary subunits (α_2 and β_3). A single exponential fit to the declining phase of the doe-1 current yielded a time constant (τ_h) of 12.0 ± 1.5 ms (mean $\pm$ s.d., $n = 4$, $V_t = 30$ mV). This decay

FIG. 1 Expression of doe-1 current in *Xenopus* oocytes. a, Schematic representation of the non-L-type subfamily of calcium channel α_1 subunits, indicating that doe-1 and the mammalian BII channel form a distinct branch¹⁶ from the branch represented by BI, rbB and doe-4. Percentage of amino-acid identity is indicated. b, Representative traces of doe-1 currents evoked from a holding potential of -80 mV to various test potentials in 10 mV increments. The α_1 subunit of doe-1 was coexpressed with an α_2 subunit derived from rabbit skeletal muscle¹¹ and a β_3 subunit derived from rabbit heart and expressed in brain²⁵. c, $I-V$ curve of doe-1 current from the same experiment as in b. d, Steady-state inactivation curve. Smooth curve is a Boltzmann function, $h_{\infty} = [1 + \exp \{(V - V_{1/2})/k\}]^{-1}$, with $V_{1/2} = -43.1$ mV, $k = 7.28$ mV. METHODS. The plasmid pDoe-1 carrying the entire protein-coding region of the channel was constructed in the *SalI*-*XbaI* site of pBluescript SK+ from plasmids containing three overlapping cDNA clones (ME2, IIL and A1) as follows. The *BstXI* (1,388)-*BamHI*(polylinker) fragment of clone pME2 (in pBluescript SK+) was ligated into *BstXI*-*BamHI* cleaved pcDNAII(Invitrogen) to yield pME2a. The *Apal* (polylinker)-*BsmI* (1,904) fragment of pME2a was ligated into *Apal*-*BsmI* cleaved plasmid IIL (in pBluescript SK-) to obtain the clone pME2-IIL. The *EagI* fragment (1,723-polylinker) of pME2-IIL was ligated into the *EagI* site of pME2 yielding pME2-IILa. Plasmid pA1-2.3 was constructed by ligating the *SacI*(3,551)-*EcoRI*(5,620) fragment of plasmid A1 into pcDNAII. The *SacI* fragment (2,574-3,551) of plasmid IIL was ligated into pA1-2.3 to obtain pA1-2.3-IIL. The 2.7 kb *EcoRI*(5,620-polylinker) fragment of plasmid A1 was ligated into the *EcoRI* site of pA1-2.3-IIL to yield pA1-IIL. The *SalI*(2,744)-*XbaI*(polylinker) fragment of pA1-IIL was subcloned into pBluescript SK+ to obtain plasmid pA1-IIL(SK+). The *SalI*(polylinker-2,744) fragment of pME2-IILa was subcloned into pA1-IIL(SK+) to obtain pDoe-1. Complementary RNA was synthesized *in vitro* using T7 or SP6 polymerase and the pDoe-1 plasmid, or pSPCBI-2 (BI-2) or pSPCA1 (CARD) plasmids (provided by T. Tanabe and Y. Mori), and pCaB1, pCaB2b, or pCaB3 plasmids (donated by V. Flockerzi and F. Hofmann). Each subunit was injected in equimolar ratios using $0.1 \mu\text{g ml}^{-1}$ α_1 , $0.1 \mu\text{g ml}^{-1}$ α_2 , and $0.03 \mu\text{g ml}^{-1}$ β ; ~ 50 nl was injected per cell. The doe-1 currents were recorded by a standard two-microelectrode voltage-clamp using a Warner amplifier (OC-



725A). The bath solution was chloride free, containing (in mM): $\text{Ba}(\text{OH})_2$, 40; tetraethylammonium-OH, 40; KOH, 2; HEPES, 5; pH 7.4 adjusted with methanesulphonic acid. Data were sampled at 10 kHz and filtered at 2 kHz. Leak and capacitance currents were subtracted off-line by a P/4 protocol. Voltage pulses were delivered every 15 s. For steady-state inactivation experiments, voltage pulses were delivered every 20 s. Peak normalized currents were fitted to a Hodgkin-Huxley inactivation curve. The magnitude of the current was somewhat variable, but currents were often up to $2 \mu\text{A}$ in size with 5 mM Ba^{2+} as the charge carrier, far larger than endogenous currents. The currents analysed for this paper were always $>0.5 \mu\text{A}$.

is at least 10-fold more rapid than for BI or cardiac L-type channels expressed in *Xenopus* oocytes (Fig. 2a), and also considerably faster than the decay of the expressed rbB N-type channel clone¹⁴, or T, N or L-type channels in sensory neurons²³. The voltage-dependence of doe-1 inactivation (Fig. 2b) is steeper than those of previously characterized HVA Ca^{2+} channels²⁴. The inactivation rate also varied with β -subunits (Fig. 2c). The rate of doe-1 inactivation at strong depolarizations (filled bars) was fastest with β_3 , and progressively slower with β_1 and β_{2b} , respectively; the same trend was found for BI (Fig. 2c, open bars), in line with previous observations²⁵ for cardiac L-type α_1 . For any given β -subunit, inactivation of doe-1 was >10-fold faster than inactivation of BI. Clearly detectable doe-1 currents were obtained with coexpression of just a β -subunit or just α_2 . The resulting currents were kinetically distinguishable from the endogenous current, although they were smaller in magnitude than the current observed with coexpression of all three subunits¹¹. The α_1 subunit alone, however, never produced a detectable current.

Single-channel properties of doe-1 were determined in cell-attached patch recordings. Unitary current records taken at three different test potentials are shown in Fig. 3. Upon depolarization, doe-1 channels open briefly and repeatedly. Channel openings were clustered near the beginning of the test pulse, particularly with stronger depolarizations. Occasionally, channel activity was spread throughout the entire sweep (Fig. 3a). Ensemble-averaged currents obtained from macro-patch recordings (Fig. 3b) were very similar in time course to those obtained from two-electrode voltage-clamp experiments (Fig. 1b). A plot of unitary current amplitude against V_t (Fig. 3c) yielded a unitary

Ba^{2+} conductance of 14 pS, similar to that of N-type channels, but larger than T-type and smaller than L-type channels in sensory neurons²³. Very similar unitary current records were obtained with Ca^{2+} as the charge carrier (data not shown).

The doe-1 channel displays an unusual pharmacological profile. Doe-1 current was inhibited by the cone snail peptide ω -CTx-GVIA (Fig. 4a); 5 μM ω -CTx-GVIA blocked $84.3 \pm 0.6\%$ of the current (mean $\pm$ s.d., $n=3$). Unlike the potent ω -CTx-GVIA block of rbB N-type channels¹⁴, whose onset is rapid and irreversible, ω -CTx-GVIA block of doe-1 develops slowly ($\tau \sim 15$ min), and is gradually but completely reversible ($\tau \sim 40$ min). Unlike L-type channels, doe-1 was not affected by the DHP antagonists nifedipine and nimodipine (5–10 μM , data not shown); moreover, it was slightly blocked rather than enhanced by the DHP agonist Bay K 8644 (Fig. 4b). We also tested the effects of the funnel web spider toxin ω -Aga-IVA, a potent blocker²⁶ of the P-type channel of cerebellar Purkinje cells⁵, and the cone snail toxin ω -CTx-MVIIC (SNX-230), a very effective inhibitor (W. A. Sather *et al.*, manuscript in preparation) of the BI clone expressed in *Xenopus* oocytes¹¹. Doe-1 was unaffected by ω -Aga-IVA (200 nM, $n=3$) and was only slightly reduced by ω -CTx-MVIIC (5 μM , $n=3$; Fig. 4b). Like all other HVA Ca^{2+} channels^{24,27}, doe-1 was very sensitive to cadmium ions (Fig. 4a, b).

Unlike most HVA Ca^{2+} channels (refs 11 and 27, but see ref. 28), doe-1 is also sensitive to blockade by nickel ions. Application of 30 μM Ni^{2+} reduced the current by about 50% (Fig. 4c, d). Washout of Ni^{2+} was associated with prompt and full recovery (not shown). The dose-dependence of Ni^{2+} block followed a one-to-one binding curve with a half maximal inhibitory

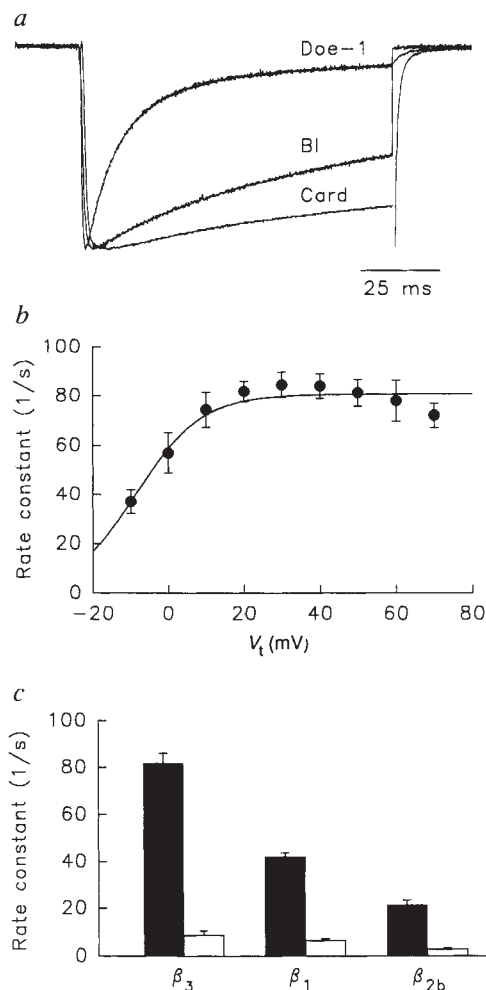


FIG. 2 Fast inactivation of doe-1 currents. *a*, Comparison between inactivation kinetics of doe-1, BI and cardiac L-type channels. The α_1 subunits were coexpressed with a fixed combination of auxiliary subunits ($\alpha_2 + \beta_3$). All currents were recorded in 40 mM Ba^{2+} at $V_t = +30$ mV. *b*, Inactivation rate constants (τ_h^{-1} , mean $\pm$ s.d., $n=4$, two batches of oocytes) plotted against test potential. τ_h was obtained by fitting the declining phase of the doe-1 current (coexpressed with α_2 and β_3 subunits) with a single decaying exponential plus a constant. Collected data for τ_h^{-1} were fit to the function: $\tau_h^{-1} (\text{s}^{-1}) = 1,000 / \{4.62 \times \exp[-V_t / (8.65 \text{ mV})] + 12.38\}$. *c*, Influence of various β -subunits on rate of inactivation (τ_h^{-1} , ordinate). The doe-1 (solid bars) or BI (open bars) α_1 subunits were coexpressed with α_2 and various β -subunits as indicated. Error bars indicate $\pm$ s.e.m. All currents were recorded with 40 mM external Ba^{2+} solution, $V_t = +20$ mV.

concentration (IC_{50}) of $33 \mu M$ (Fig. 4c). Current-voltage relationships show Ni^{2+} block at all potentials (Fig. 4d); some relief from blockade may occur at the most positive potentials.

There is considerable interest in HVA Ca^{2+} channels in mammalian neurons that fall outside the categories of L-, N- and P-type but contribute substantially to Ca^{2+} entry. Characterization of *doe-1* channels in oocytes raises the question of whether mammalian central nervous system neurons possess a similar current. Indeed, we have found that rat cerebellar granule cells exhibit a rapidly decaying, Ni^{2+} - and Cd^{2+} -sensitive HVA current (Fig. 4e). Isolated by blockade of other channel types (Fig. 4 legend), this transient current is high-voltage activated; peak current increases steeply over the range between -40 and 0 mV (Fig. 4f). In 5 mM Ba^{2+} at 0 mV, the peak amplitude of the Ba^{2+} current was 58.5 ± 3.1 pA ($n=97$), or $\sim 20\%$ of the total Ba^{2+} current. Inactivation is rapid ($\tau_h = 22.2 \pm 2.3$ ms at 0 mV, 23.3 ± 3.2 ms at $+30$ mV, $n=5$). Additionally, steady-state inactivation occurs at relatively negative potentials ($V_{1/2} = -62$ mV in 5 mM Ba^{2+}). These characteristics are similar, although not identical with those of *doe-1*. The pharmacology of this cerebellar granule cell current is also largely similar to that of *doe-1*. It appears resistant to nimodipine ($10 \mu M$), ω -CTx-GVIA (1 – $5 \mu M$), ω -CTx-MV1IC ($5 \mu M$) and ω -Aga-IVA (100 nM). For example, in 30 – 100 nM ω -Aga-IVA the peak current amplitude at 0 mV was 57.2 ± 5.4 pA ($n=15$), virtually the same as in the absence of the agatoxin. As with *doe-1*, the granule cell current is sensitive to Ni^{2+} ($IC_{50} = 66 \mu M$, Fig. 4c), and like other HVA currents, it is also blocked by Cd^{2+} ($51 \pm 3\%$ block by $1 \mu M$ Cd^{2+} , $n=6$). One difference, however, is that the mammalian current is not affected by ω -CTx-GVIA even at concentrations up to $5 \mu M$ ($n=5$). Overall, biophysical and pharmacological comparisons suggest that the granule cell current could be a mammalian counterpart to the *doe-1* channel. The BII clone, whose transcripts appear in the cerebellum and other brain regions¹⁵, is the mammalian channel most homologous to *doe-1*.

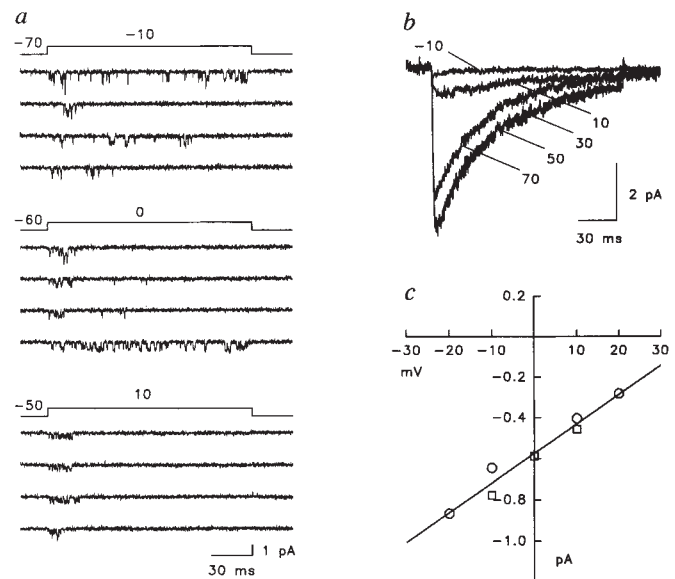


FIG. 3 Behaviour of single *doe-1* channels. *a*, Unitary current records obtained at different membrane potentials as indicated. *doe-1* coexpressed with α_2 and β_1 . *b*, Ensemble-averaged currents obtained from a macro-patch. V_h was -70 mV and V_t ranged from -10 to $+70$ mV as marked. *c*, Single-channel amplitude plotted against V_t . Data from two separate experiments (circles, squares) were fitted to a linear function. The slope of the line corresponds to a single-channel conductance of 14.4 pS.

METHODS. Single-channel records of *doe-1* were obtained from cell-attached patches using either a List EPC-7 or an Axon 200A amplifier. The membrane potential of oocytes was zeroed with a solution containing (in mM): K-aspartate, 100 ; EGTA, 10 ; HEPES, 10 ; pH 7.4 with KOH. The pipette solution contains (in mM): $BaCl_2$, 115 ; HEPES, 10 ; pH 7.4 . Data were sampled at 10 kHz and filtered at 2 kHz. The single channel amplitude was obtained by fitting the amplitude histogram to a gaussian distribution.

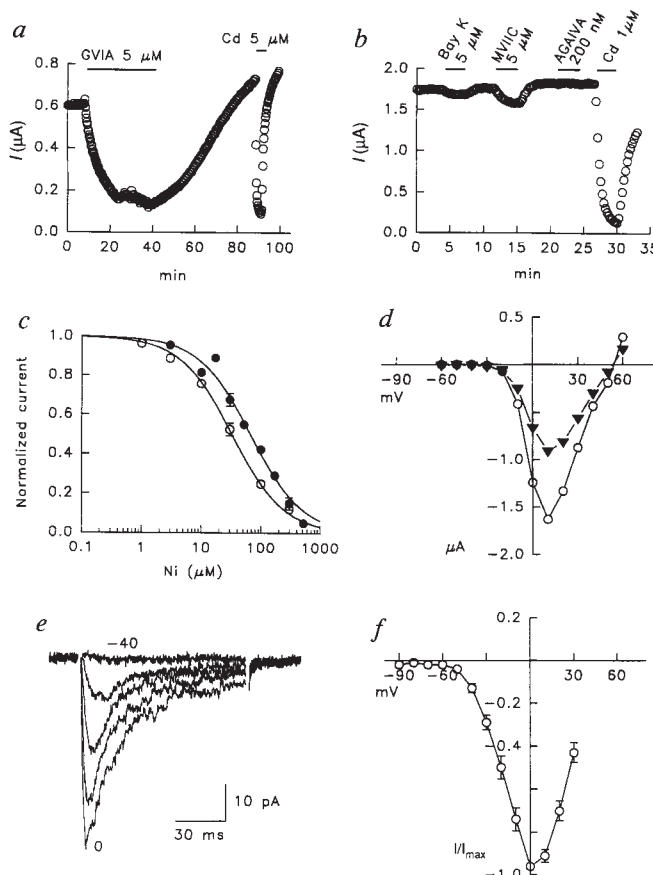


FIG. 4 Pharmacological properties of *doe-1* Ca^{2+} channels and similar channels in rat cerebellar granule neurons. To minimize the effects of divalent ions on toxin block, 5 mM external Ba^{2+} was used for these experiments. *a*, Effect of ω -CTx-GVIA ($5 \mu M$) and Cd^{2+} ($5 \mu M$) on *doe-1* current magnitude. *b*, Effects of a series of pharmacological agents, including Bay K 8644 ($5 \mu M$), ω -CTx-MV1IC (SNX-230) ($5 \mu M$), ω -Aga-IVA (200 nM) and Cd^{2+} ($1 \mu M$). *c*, Concentration-dependence of Ni^{2+} block of *doe-1* current ($\circ$, $n=4$) and rapidly decaying cerebellar granule cell current ($\bullet$, $n=2$ – 11 , 27 cells). The smooth curves are one-to-one binding curves with IC_{50} s of $32.5 \mu M$ ($\circ$) and $66.4 \mu M$ ($\bullet$). *d*, Peak current-voltage relationship for *doe-1* remains largely unchanged in the presence of $30 \mu M$ Ni^{2+} . *e*, Transient Ba^{2+} current recorded from a cerebellar granule cell neuron in the presence of nimodipine ($10 \mu M$), ω -CTx-GVIA ($1 \mu M$), ω -CTx-MV1IC ($5 \mu M$), ω -Aga-IVA (30 – 100 nM) had no significant effect ($n=16$, not shown). *f*, Peak current-voltage relationship. Collected data from 13 cells, normalized by largest peak current (bars give s.e.m.).

METHODS. Cerebella were removed from the brains of p4–p5 rats, briefly trypsinized (4 min, 10 mg ml^{-1}), and dissociated in Ca^{2+} -free medium. Dissociated cells were plated on matrigel-coated coverslips, incubated for 5 – 11 days in a standard serum-supplemented culture medium before use. Whole-cell recordings were made with 2 – 8 Mohm patch clamp pipettes filled with (in mM): Cs-methylsulphate, 108 ; $MgCl_2$, 4.5 ; EGTA-CsOH, 9 ; ATP, 4 ; GTP, 0.3 ; HEPES-CsOH, 24 (pH 7.3 , 298 mOsm). Ba^{2+} currents were isolated in a bath solution containing (in mM) TEA-Cl, 160 ; $BaCl_2$, 5 ; HEPES, 10 ; TTX, 0.001 ; cytochrome *c*, 0.008 (pH 7.3 , 313 mOsm). Seal resistances were >5 Gohm, series resistance >20 Mohm (typically 10 Mohm), and membrane capacitance about 5 pF. Drugs were applied using a fast application system, full solution exchange being achieved in <1 s. Ba^{2+} or Ca^{2+} currents were evoked by stepping from a holding potential of -80 to the test potential for 105 ms every 10 s. The recorded currents were filtered (2 kHz) and leak-subtracted.

(Fig. 1a). Thus BII may encode the granule cell current.

The functional properties of *doe-1* present an interesting contrast with aspects of previously characterized Ca^{2+} channels. Although these channels inactivate rapidly and are quite sensitive to Ni^{2+} , like T-type channels^{24,29}, they differ markedly from T-type channels with regard to their voltage-dependence and their sensitivity to Cd^{2+} . The channels described here differ from L-type and P-type channels by virtue of their kinetics and their insensitivity to DHPs and ω -Aga-IVA, respectively. Thus, *doe-1* and its putative mammalian counterpart can be readily distinguished from T-, L- and P-type channels.

Doe-1 and the granule cell current share more functional properties with N-type channels than with the other channel types. But these channels are distinguished from *rbB*, a previously expressed N-type Ca^{2+} channel¹⁴, by the rapidity of their inactivation and their susceptibility to Ni^{2+} block. At least some of these distinguishing characteristics have already been reported for currents broadly classified as N-type in mammalian neuronal membranes. Rapid inactivation ($\tau \sim 10\text{--}20\text{ ms}^{30}$) is an earmark of the N_1 channel of peptide-releasing mammalian nerve terminals^{30,31}. A component of N-type current that is reversibly blocked by ω -CTx-GVIA has been described in peripheral neurons³². Together, these results highlight the diversity of Ca^{2+} channels that have been or could be considered within the general category of N-type. Some of these channels may be found to arise from *doe-1* or BII-like clones, rather than the *rbB* gene.

Until now, functional classes of Ca^{2+} channels seemed to map neatly onto a structural family tree, with separate limbs for N-type, P-type and various L-type channels^{2,6}. In this context, the resemblance of *doe-1* to various N-type channels is particularly noteworthy because there is less structural homology between *doe-1* and *rbB* (clearly N-type¹⁴) than between *rbB* and BI (clearly not N-type¹¹). Thus, Ca^{2+} channel classifications based on kinetics and pharmacology can cut across categories based on sequence homology. A similar principle is emerging for K^{+} channels³³. Structurally, *doe-1* is most akin to the BII channel of mammalian brain¹⁵ (68% identity). Once BII has been expressed, it will be interesting to compare it to *doe-1* and the cerebellar granule cell current.

The rapid inactivation of *doe-1*-like channels may have important consequences for neuronal function. Such inactivation would strongly influence the ability of a channel to contribute to Ca^{2+} entry during repetitive depolarizations. This might reduce the activation of Ca^{2+} -activated K^{+} channels and thereby preserve the excitability of the neuron. At a nerve terminal, such properties could decrease synaptic efficacy during a train of impulses. Our results with *doe-1* emphasize the extent to which Ca^{2+} channels have diversified to allow neuronal responses to be tuned to particular patterns of electrical activity. □

25. Hulin, R. *et al.* *EMBO J.* **11**, 885–890 (1992).

26. Mintz, I. *et al.* *Nature* **355**, 827–829 (1992).

27. Kasai, H. & Neher, E. *J. Physiol., Lond.* **448**, 161–188 (1992).

28. Regan, L. J. *J. Neurosci.* **11**, 2259–2269 (1991).

29. Narahashi, A. *et al.* *J. Physiol., Lond.* **383**, 231–249 (1987).

30. Lemos, J. R. & Nowycky, M. C. *Neuron* **2**, 1419–1426 (1989).

31. Wang, X. *et al.* *J. Physiol., Lond.* **445**, 181–199 (1992).

32. Plummer, M. R. *et al.* *Neuron* **2**, 1453–1463 (1989).

33. Salkoff, L. *et al.* *Trends Neurosci.* **15**, 161–166 (1992).

ACKNOWLEDGEMENTS. We thank C. Jung and J. Phillips for technical assistance in constructing pDoe-1. F. Schweizer and B. Wendland for preparation of cerebellar granule cells, G. Miljanich (Neurex) for providing ω -CTx-MVHC (SNX-230), M. Adams (UC Riverside) for a gift of ω -Aga-IVA and R. W. Aldrich, D. D. Friel, H. Reuter and W. A. Sather for comments on the manuscript. This work was supported by the NIH (T.L.S., R.W.T.), the Wellcome Trust (A.D.R.), a fellowship from the American Heart Association (California Affiliate) (J.-F.Z.) and gifts from Miles, Parke-Davis, Wyeth-Ayerst and Eli Lilly (R.W.T.). A.D.R. is a Lucille P. Markey Visiting Fellow.

Germ-line mutations of the *RET* proto-oncogene in multiple endocrine neoplasia type 2A

Lois M. Mulligan*, John B. J. Kwok*, Catherine S. Healey*, Mark J. Elsdon*, Charis Eng*, Emily Gardner*, Donald R. Love†, Sara E. Mole†, Julie K. Moore*, Laura Papi†, Margaret A. Ponder*, Hakan Telenius*, Alan Tunnacliffe* & Bruce A. J. Ponder*‡

Cancer Research Campaign Human Cancer Genetics Group, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

MULTIPLE endocrine neoplasia type 2A (MEN 2A) is a dominantly inherited cancer syndrome that affects tissues derived from neural ectoderm. It is characterized by medullary thyroid carcinoma (MTC) and pheochromocytoma¹. The *MEN2A* gene has recently been localized by a combination of genetic and physical mapping techniques to a 480-kilobase region in chromosome 10q11.2 (refs 2,3). The DNA segment encompasses the *RET* proto-oncogene, a receptor tyrosine kinase gene expressed in MTC and pheochromocytoma and at lower levels in normal human thyroid⁴. This suggested *RET* as a candidate for the *MEN2A* gene. We have identified missense mutations of the *RET* proto-oncogene in 20 of 23 apparently distinct MEN 2A families, but not in 23 normal controls. Further, 19 of these 20 mutations affect the same conserved cysteine residue at the boundary of the *RET* extracellular and transmembrane domains.

We searched for gross alterations of the *RET* locus, using pulsed field and standard Southern blots hybridized with a *RET* complementary DNA⁵ and probes derived from a *RET* cosmid contiguous sequence (contig; J.B.J.K., manuscript in preparation) but detected no altered DNA fragments in a panel of 50 unrelated MEN 2A patients (data not shown). To search for smaller mutations, we designed polymerase chain reaction (PCR) primers for an overlapping set of 16 amplicons spanning the entire *RET* coding sequence^{6,7}. Because *RET* transcripts are not detected in lymphocytes, we screened RNA from four MTCs and three pheochromocytomas from six MEN 2A patients and from a sporadic MTC and pheochromocytoma, all of which were known to express *RET* transcripts. First-strand cDNA was prepared, PCR-amplified for each amplicon and the products analysed for sequence variations by the chemical cleavage mismatch procedure (CCM)^{8,9}.

Analysis of the entire *RET* coding sequence in these nine tumour samples identified CCM variants in three regions of the

Received 30 March; accepted 20 April 1993.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1992).
- Tsien, R. W. *et al.* *Trends Pharmacol. Sci.* **12**, 349–354 (1991).
- Bean, B. P. A. *Rev. Physiol.* **51**, 367–384 (1989).
- Miller, R. J. & Fox, A. P. in *Intracellular Calcium Regulation* 97–138 (Wiley-Liss, New York, 1990).
- Llinas, R. *et al.* *Trends Neurosci.* **15**, 351–355 (1992).
- Snutch, T. P. & Reiner, P. B. *Curr. Opin. Neurobiol.* **2**, 247–253 (1992).
- Tanabe, T. *et al.* *Nature* **328**, 313–318 (1987).
- Mikami, A. *et al.* *Nature* **340**, 230–233 (1989).
- Williams, M. E. *et al.* *Neuron* **8**, 71–84 (1992).
- Hui, A. *et al.* *Neuron* **7**, 35–44 (1991).
- Mori, Y. *et al.* *Nature* **350**, 398–402 (1991).
- Starr, T. V. B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 5621–5625 (1991).
- Dubel, S. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **89**, 5058–5062 (1992).
- Williams, M. E. *et al.* *Science* **257**, 389–395 (1992).
- Niidome, T. *et al.* *FEBS Lett.* **308**, 7–13 (1992).
- Horne, W. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **90**, 3787–3791 (1993).
- Campbell, K. P. *et al.* *Trends Neurosci.* **11**, 425–430 (1988).
- Catterall, W. A. *Science* **242**, 50–61 (1988).
- Ahmad, S. N. & Miljanich, G. *Brain Res.* **453**, 247–256 (1988).
- Horne, W. A., Hawrot, E. & Tsien, R. W. *J. Biol. Chem.* **266**, 13719–13725 (1991).
- Carlson, S. S., Wagner, J. A. & Kelly, R. B. *Biochemistry* **17**, 1188–1199 (1978).
- Ngeese, J. K. *et al.* *J. Biol. Chem.* **266**, 2625–2680 (1991).
- Fox, A. P. *et al.* *J. Physiol., Lond.* **394**, 173–200 (1987).
- Fox, A. P. *et al.* *J. Physiol., Lond.* **394**, 149–172 (1987).

*Present addresses: School of Biological Sciences, University of Auckland, Auckland, New Zealand (D.R.L.); Department of Paediatrics, University College of Middlesex School of Medicine, University College London, London, UK (S.E.M.); Dip. Fisiopatologia Clinica, Sezione di Genetica Medica, Viale Pieraccini 6, Firenze, Italy (L.P.).

‡To whom correspondence should be addressed.

TABLE 1 *RET* mutations in MEN 2A

Codon	Position* Base pair	Family	Base pair changes	Amino-acid changes
364	1,783	A34	TGC→GGC	Cys→Gly
378-380	1,827-1,831	A33†, M01 M04, M09, M10 M11, M14, M19 M32, A12, 165, 167	GAG CTG TGC→GAC GTG CGC	Glu Leu Cys→Asp Val Arg
380	1,831	123, M18, F193	TGC→GGC	Cys→Gly
	1,832	M06, F060 A11 M07	TGC→TAC TGC→TCC TGC→TTC	Cys→Tyr Cys→Ser Cys→Phe

* Amino-acid and codon position refer to the *ret* sequence of ref. 6.

† Cys 380 sequence was investigated in tumour material only for this family.

gene (Fig. 1). PCR products identified by cleavage of variant sequence were isolated and sequenced directly. Two of these variants represented conservative base-pair substitutions at positions 2,238 and 2,643 (numbered according to the published sequence of ref. 6) in the tyrosine kinase domain and are probably normal sequence polymorphisms (E.G. *et al.*, manuscript in preparation).

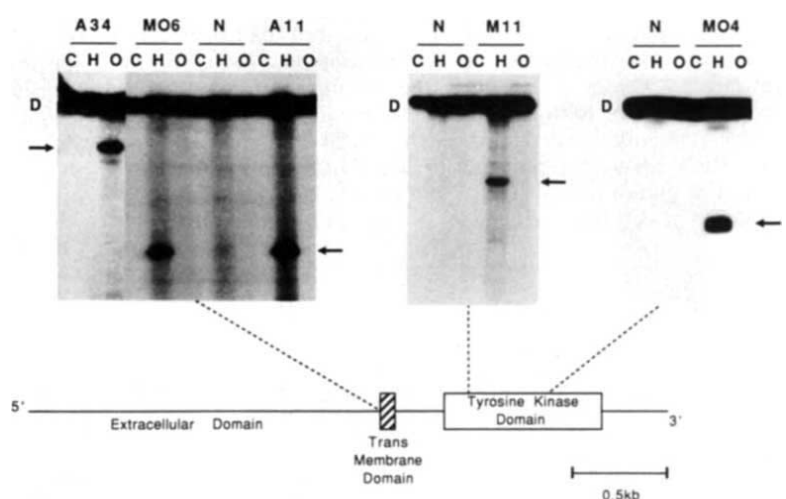
The remaining CCM variants were detected in a 413-base-pair (bp) PCR amplicon spanning the *RET* transmembrane domain and 5' portion of the tyrosine kinase domain (Fig. 1). Sequence analysis of these PCR products identified an altered sequence in all seven tumour samples from six patients with MEN 2A (Table 1; A12, A33, A34, 167, M04, M11). The samples from the two sporadic tumours had no sequence alterations. Six of the MEN 2A samples, including an MTC and pheochromocytoma from the same individual, proved to have an identical, three-base sequence change (GAGCTGTGC→GACGTGCGC) resulting in the substitution of three amino acids in codons 378-380 (Glu-Leu-Cys→Asp-Val-Arg). The presence of all three changes on a single allele was confirmed by sequencing cloned *RET* cDNAs from an MEN 2A pheochromocytoma cDNA

library (Fig. 2a). The library was prepared from an individual (A12) heterozygous for the mutant allele. The codon for cysteine at position 380 (Cys 380) is one of 27 cysteine residues in the *RET* extracellular domain that are conserved between man and mouse^{6,7,10}. In the remaining MEN 2A tumour sample (A34) there was no mutation in Cys 380. But a T→G conversion on codon 364 again led to replacement of a conserved cysteine residue, in this case by a glycine. In all but one tumour, heterozygosity of the DNA sequence indicated that both normal and mutant *RET* alleles were expressed. The exception was in the MTC of individual 167-302, which expressed only the mutant *RET* allele (Table 1).

To determine whether the *RET* mutations found in the tumours were germ-line or somatic in origin, we used PCR primers derived from *RET* intronic sequences to analyse normal tissue DNA from 21 unrelated MEN 2A individuals, including 4 of the 6 MEN 2A individuals with tumour mutations and from 23 normal controls (Fig. 2b). *RET* mutations of codon 380 were detected in 17/21 germ-line MEN 2A samples and the codon 364 mutation was detected in one (A34). Normal *RET* sequence was observed in all normal controls. In every case in which a *RET*

FIG. 1 Chemical cleavage mismatch analysis (CCM) of *RET*. Autoradiograms showing variations in CCM cleavage pattern of PCR-amplified *RET* transcripts. MEN 2A family members are indicated together with normal (N) controls. C, Control lanes containing uncleaved heteroduplex. H and O, Hydroxylamine- and osmium tetroxide-modified heteroduplexes, respectively, cleaved by treatment with piperidine. The positions of the heteroduplex is marked by D and the *RET* cleavage products by arrows. The positions of the cleaved bases relative to the *RET* extracellular, transmembrane and tyrosine kinase domains are shown.

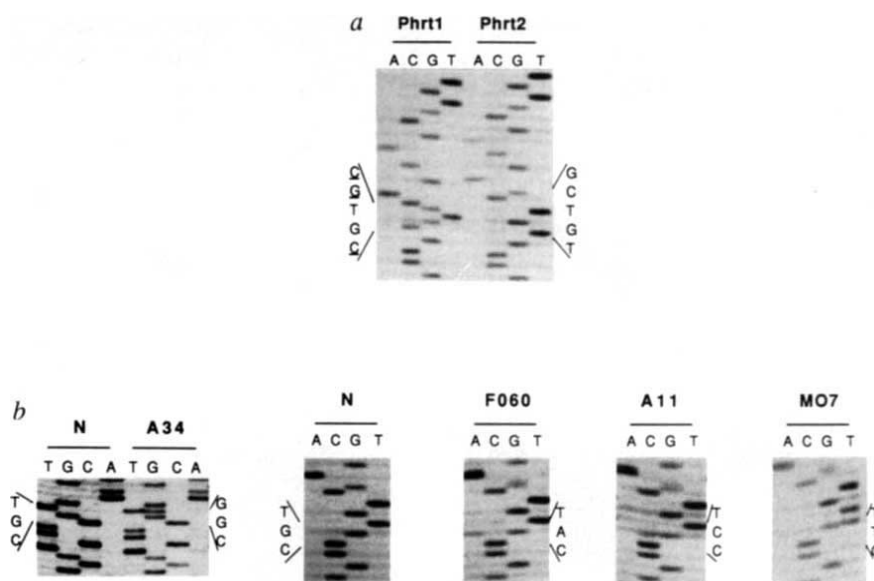
METHODS. PCR primers used in analysis of *RET*. Transmembrane domain (codons 262 to 399): CRT2S (5'CCCGTTCCTGTCAGTCAGCAAG3') and CRT2A (5'GACAGCACCGAGACGAT3'); 5' tyrosine kinase domain (codons 480 to 604): CRT4S (5'GAAAAAGTGGTCAAGGC-AAC3') and CRT4A (5'GATCTGCCAGGCAAATGAGATG3'); 3' tyrosine kinase domain (codons 590 to 716): CRT5S (5'GGGCCCTCACCATGGGCGACCT3') and CRT5A (5'CCA-TCCGGTGGCCGGTCTTCAAG3'). PCR amplification of poly(T)-primed first-strand cDNA was done for 55 cycles of 1 min each at 95 °C, and 55 °C and 72 °C. A normal control fragment was gel-purified in low-melting-point agarose, end-labelled using [³²P]-γ-ATP and T4 polynucleotide kinase and hybridized to the corresponding PCR fragment from each sample. The heteroduplex was modified by treatment with either 2.5 M hydroxylamine or



0.025% osmium tetroxide and modified bases cleaved by addition of 10% piperidine. The cleavage products produced by sequence variations were visualized by separation on 5% polyacrylamide gels and autoradiography.

FIG. 2 Sequence of codons 378 to 380 of the *RET* proto-oncogene in MEN 2A individuals. *a*, Sequence of the 378 to 380 codons in two copies of *RET* isolated from a pheochromocytoma cDNA library prepared from patient A12 who is heterozygous for the 378–380 *RET* mutation. PHRT2 contains the normal *RET* sequence. PHRT1 has three base-pair changes (underlined) showing that all three changes are present in a single *RET* allele. *b* Germ line sequence of normal individuals (N) and MEN 2A patients from families A34, F060, A11 and M07. The mutant sequence for the relevant codons is shown to the right of each sample. The *RET* normal DNA sequence is shown to the left.

METHODS. PCR primers used were CRT2B (5'CGTATGCGTCGG3') and CRT2A. PCR amplification and product purification were as described in the legend of Fig. 1. Amplification products were sequenced using the CRT2A primer with the Δ TaqI sequencing kit (US Biochemicals). Plasmid DNAs were sequenced directly with the Sequenase kit (US Biochemicals).



mutation had been observed in tumour samples, the same sequence variant was present in the germ-line, indicating that the mutations did not arise somatically during tumorigenesis.

If *RET* mutations are responsible for the MEN 2A phenotype, the mutant alleles should segregate with the disease in MEN 2A families. Confirmation of this was obtained in 8 families (M04, M06, M07, M11, M14, 167, M18, 123) which had the same mutation of codon 380 in two or more affected family members but not in unaffected family members, confirming that the mutation was on the disease chromosome in each case.

In total, 19 distinct mutations at codon 380 were detected, representing 5 different amino-acid changes (Table 1). Despite sharing the same *RET* codon 380 mutation, some families (M09 and M19; 123 and M18) did not share haplotypes for other polymorphisms^{11–13} within the *RET* locus (<30 kb away; J.B.J.K. *et al.*, manuscript in preparation). This suggests that at least some of these similar mutant alleles have arisen by independent mutation.

The MEN 2B syndrome, similar to MEN 2A but characterized by earlier tumour onset, ganglioneuromatosis of the intestine and Marfanoid habitus, also maps to the 10q11.2 region². We did not find mutation of codon 380 in any of 10 MEN 2B patients examined. This may not be surprising, considering the differences in disease phenotype. We cannot, as yet, predict whether MEN 2B is due to a different mutation in the *RET* gene or to mutation of a closely linked gene.

Taken together, our data show that 20/23 unrelated patients with MEN 2A are heterozygous for *RET* missense mutations in germ-line and/or tumour DNA. In 19 out of 20 cases (95%) the mutation affects Cys 380. The *RET* cysteine residues are highly

conserved between species and 27 of 28 cysteines, including Cys 380, are present in the mouse *RET* extracellular domain^{6,7,10}. These data suggest that replacement of the Cys 380 residue may disrupt normal *RET* protein conformation, altering *RET* signal transduction and leading to the MEN 2A phenotype. The remaining three MEN 2A patients do not have sequence changes in codon 380, however, so it seems likely that mutation of positions elsewhere in the *RET* extracellular domain may be responsible for MEN 2A in these cases.

The *RET* gene is structurally rearranged in 25% of papillary thyroid carcinomas by juxtaposition of *RET* transmembrane and tyrosine kinase domains with other 5' expressed sequences¹⁴. Clearly, similar activating mechanisms do not occur in MEN 2A. If replacement of Cys 380 and heterozygous expression of the mutant *RET* isoform is responsible for MEN 2A, a dominant or dominant-negative mechanism for the disease seems more probable than the classic tumour suppressor model. This is consistent with the absence of allele loss reported for chromosome 10q in MTC and pheochromocytoma^{15–17}. The specificity of the *RET* mutants observed in Cys 380 is reminiscent of the codon specificity of oncogenic point mutations of other tyrosine kinase receptor genes such as *c-fms*, *neu* and *trk*^{18–20}. *In vitro* analyses have shown that substitution of specific amino acids in the extracellular (*fms*, *trk*^{18–19} or transmembrane (*neu*²⁰) domains of these proteins can induce cellular transformation. If the *RET* Cys 380 mutations observed in MEN 2A are functionally analogous, then this would be the first example, to our knowledge, of dominantly acting oncogenic point mutations as the initiating events in human hereditary neoplasia. □

Received 18 March; accepted 15 April 1993.

- Schimke, R. N. *Am. Rev. Med.* **35**, 25–31 (1984).
- Gardner, E. *et al. Hum. molec. Genet.* **2**, 241–246 (1993).
- Mole, S. E., Mulligan, L. M., Healey, S. C., Ponder, B. A. J. & Tunncliffe, A. *Hum. molec. Genet.* **2**, 247–252 (1993).
- Santoro, M. *et al. Oncogene* **5**, 1595–1598 (1990).
- Ishizaka, T. *et al. Oncogene* **4**, 1519–1521 (1989).
- Takahashi, M. *et al. Oncogene* **3**, 571–578 (1988).
- Takahashi, M., Buma, Y., Hiai, H. *Oncogene* **4**, 805–806 (1989).
- Cotton, R. G., Rodrigues, N. R. & Campbell, R. D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4397–4401 (1988).
- Rodrigues, N. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 7555–7559 (1990).
- Iwamoto, T. *et al. Oncogene* **8**, 1087–1091 (1993).
- Mulligan, L. M., Gardner, E., Papi, L. & Ponder, B. A. J. *Nucleic Acids Res.* **19**, 5795 (1991).
- Yamamoto, M. *et al. Jap. J. clin. Oncol.* **21**, 149–152 (1991).
- Tahira, T. *et al. Nucleic Acids Res.* **18**, 7472 (1991).
- Fusco, A. *et al. Nature* **328**, 170–172 (1987).

- Mulligan, L. M., Gardner, E., Smith, B. A., Mathew, C. G. P. & Ponder, B. A. J. *Genes Chrom. Cancer* **6**, 166–177 (1993).
- Landsvater, R. M. *et al. Genomics* **4**, 246–250 (1989).
- Neikins, B. D. *et al. Cancer Res.* **49**, 4114–4119 (1989).
- Roussel, M. F., Downing, J. R., Rettenmier, C. W. & Sherr, C. J. *Cell* **55**, 979–988 (1988).
- Coulien, F. C. *et al. Molec. cell. Biol.* **10**, 4202–4210 (1990).
- Bargmann, C. I., Hung, C. & Weinberg, R. A. *Cell* **45**, 649–657 (1986).

ACKNOWLEDGEMENTS. We thank the MEN2 families and the many clinicians in the CRC Medullary Thyroid Cancer group and in the MEN2 International Collaborative Group who have provided us with clinical material and, in particular, C. Calmettes and G. Lenoir at the GETC group, A. Frilling, H. Neumann, S. Reichlin, M. Telenius-Berg and T. Cummings; M. Jackson, R. Eeles and N. Rodrigues for discussion and P. Harrington for technical assistance. This study was supported by grants from the EEC (L.P.), Swedish Medical Council (H.T.), A.O. Swards Fund (H.T.), M. Brogström's Fund (H.T.), CRC-Dana Farber Cancer Institute Fellowship (C.E.) and by a programme grant from the Cancer Research Campaign. B.A.J.P. is a Gibb Fellow of the CRC.

kidney. The results are consistent with immunohistological studies showing high expression of MAD in HEV in PP, MLN, some expression in the marginal sinus around splenic white pulp nodules in the spleen¹², and focal weak reactivity in PLN HEV in mice of this age.

We examined adhesion of the mucosal HEV binding lymphoma TK1¹ to the expressed addressin. COS cells transfected with full-length cDNA in the CDM8 expression vector¹³ stain specifically with antibodies MECA-367 and MECA-89, against two distinct epitopes on Mad⁴ (data not shown). TK1 cells bind to addressin-transfected COS cells and not to control (mock) transfectants (Fig. 2*b, d*). Binding is specific to cells expressing MAD as shown by simultaneous immunofluorescence staining with the nonblocking monoclonal antibody, MECA-89 (ref. 4; 2*c*). Adhesion is specifically blocked by anti-addressin antibody MECA-367 (Fig. 3*c*). Furthermore, binding is specific for the mucosal HEV binding cell line as L1-2 (ref. 5) and NS8 (ref. 14) cells (lines previously shown not to adhere to the addressin) fail to adhere to the COS cell transfectants.

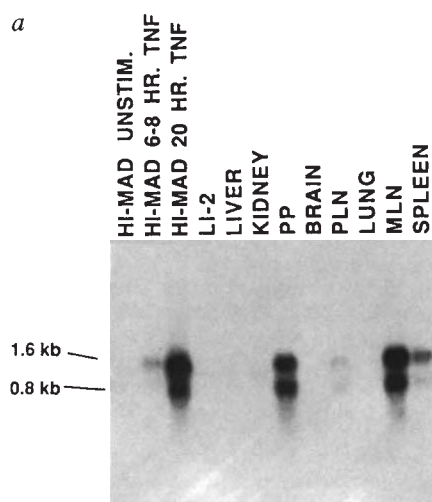
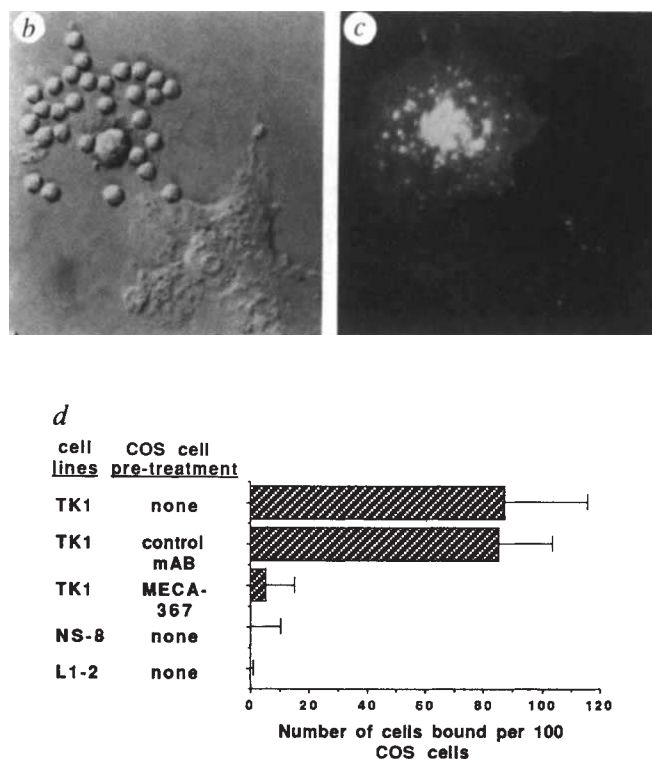


FIG. 2 *a*, Expression of MADCAM-1 cDNA. Poly(A)⁺ RNA from unstimulated hiMAd-4 cells (lane 1), hiMAd-4 cells stimulated with TNF- α for 6–8 h (lane 2), hiMAd-4 cells stimulated with TNF- α for 20 h (lane 3), the L1-2 pre-B lymphoma (lane 4), normal 8-week BALB/c mouse liver (lane 5), kidney (lane 6), Peyer's patches (lane 7), brain (lane 8), peripheral lymph nodes (lane 9), lung (lane 10), mesenteric lymph nodes (lane 11) and spleen (lane 12) were hybridized to the 1.4-kb MADCAM-1 cDNA insert. *b–d*, Functional analysis of the transfected MADCAM-1 cDNA. *b*, Adherence of TK1 lymphoma cells to COS cells transfected with MADCAM-1 cDNA, viewed with Hoffman modulation. *c*, The same field as in *b* was stained with anti-mucosal addressin antibody MECA-89 and viewed under ultraviolet illumination. *d*, Adhesion to MADCAM-1 COS cell transfectants is selective for mucosal homing lymphomas and is specifically inhibited by anti-mucosal addressin antibody MECA-367.

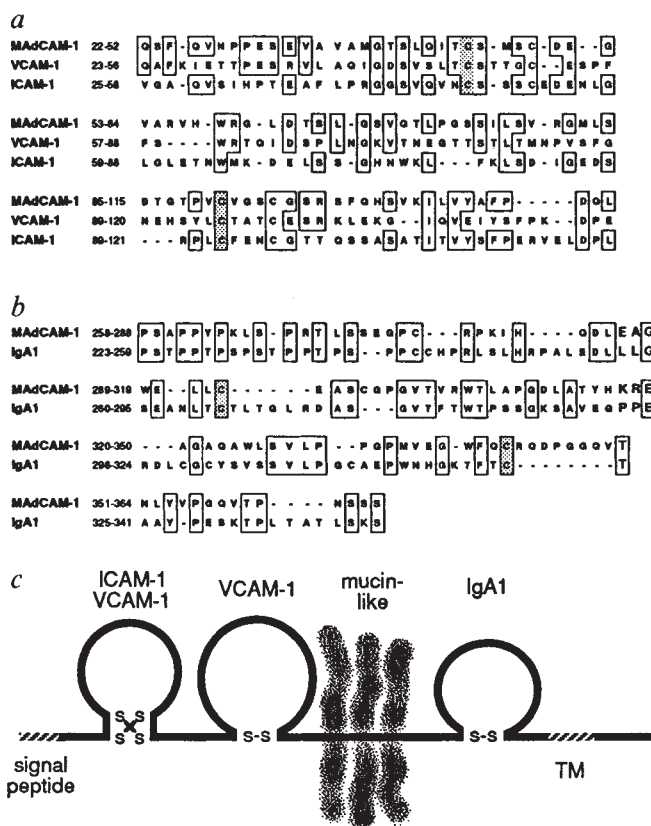
METHODS. *a*, Northern blots used 1 μ g of poly(A)⁺ RNA from cell lines and 2 μ g from tissues. RNA was denatured and electrophoresed through a 1% agarose formaldehyde gel and transferred to a PVDF (Immobilon, Millipore) membrane by standard capillary blot procedures²⁶. RNA samples were electrophoresed on a separate gel and stained with ethidium bromide to ensure that the quality and quantity of each cell or tissue RNA was equivalent (not shown). Hybridization was done at 65 °C using standard conditions²⁶ and blots were washed at high stringency (65 °C, 0.1 $\times$ SSC, 0.2% SDS). cDNAs were labelled with [α -³²P]dCTP by priming with random hexamers²⁶. *b–d*, The expression construct pCDMAd-15 was generated by subcloning the *SalI*-*NotI* insert of λ MAd-15 into *XhoI*-*NotI* digested pCDM8¹³. This construct was transfected onto COS cell monolayers using DEAE dextran¹³. After 48 h, cells were replated into 3.5-cm dishes to ensure even distribution of all transfectants and 24 h later transfectants were analysed by immunofluorescence. Both MADCAM-1 and mock transfectants (receiving no DNA) were stained with anti-addressin monoclonals MECA-367 or MECA-89, or with the isotype matched control MEL-14²⁷. Staining was detected both

The predicted extracellular domain of the addressin comprises three immunoglobulin-like domains, each containing the invariant cysteine residues known to stabilize the basic immunoglobulin loop^{15,16}, with distances of 45, 67 and 47 amino acids, respectively (Figs 1 and 3*c*). Because most immunoglobulin family adhesion receptors have been given names including the abbreviation 'CAM', we now refer to this mucosal addressin as MADCAM-1 (for mucosal addressin cell adhesion molecule-1). Search of the NBRF database using the FASTP program¹⁷ revealed a striking homology of the most N-terminal domain (domain one) with the first domains of rat ICAM-1⁶ (Fig. 3*a, c*) and of human VCAM-1 (Fig. 3*a, c*). The first domain in all three molecules is similar in that the spacing between the cysteine residues and the predicted β -strand structure is typical of H or C2 immunoglobulin domains^{15,16} and all three molecules have a similar arrangement of double cysteine residues at both ends of the domain. The N-terminal immunoglobulin domains of the vascular adhesion receptors ICAM-1 and VCAM-1 have been shown to contain binding sites for their respective lym-



visually and by FACS analysis (not shown) using a PE-conjugated goat Fab'2 anti-rat IgG (Caltag). Transfection efficiencies were ~25% for each experiment. For adhesion assays, COS cell transfectants were plated into 3.5-cm plates as described above. Lymphoid cells were pelleted and resuspended at a density of 2×10^6 ml⁻¹ in RPMI medium without sodium bicarbonate with 2% serum and buffered with 10 mM HEPES, pH 6.9. Cell suspensions were layered (1 ml per plate) onto the COS cell transfectants and gently rotated on a gyrotory shaker (New Brunswick Scientific model G-2) at 60 r.p.m. for 30 min at 25 °C and then washed four times in the same media by aspirating and gently filling each plate with 1 ml per wash. Plates were fixed in cold (4 °C) Dulbecco's PBS containing 1% glutaraldehyde and sodium azide. For antibody inhibition assays, transfectants were pretreated with 20 μ g ml⁻¹ anti-addressin antibody MECA-367 or isotype matched control Hermes-1²⁸ (anti-human CD44, reactive with COS cells) at 1 ml per plate for 20 min at 25 °C. Solutions were then aspirated and adhesion assays were done as described above. The number of transfectants and bound lymphoid cells in four random fields was counted. Data are presented as the number of cells bound per 100 COS cells with the background (defined as the number of lymphoid cells bound to untreated mock transfectants) subtracted $\pm$ standard error. The graph above is representative of four independent experiments with similar results.

FIG. 3 Alignment of MAdCAM-1 with homologous sequences. Two or more identical residues are boxed. Proteins were initially aligned by searching NBRF databases using the FASTP program¹⁷ and completed by inspection. Cysteines proposed to contribute to standard immunoglobulin loop disulphide bonds are shaded. *a*, Alignment of the first immunoglobulin domain of MAdCAM-1 with the first immunoglobulin domain of rat ICAM-1 and human VCAM-1. The residue identity for this alignment is 32% for ICAM-1 and 28% for VCAM-1 while significance scores are 5.91 and 3.28, respectively. *b*, Alignment of the third immunoglobulin domain of MAdCAM-1 with the $\alpha 2$ exon of gorilla IgA1. A proline/serine-rich region proximal to the normal boundary of an immunoglobulin domain is included in the alignment because the hinge region of IgA1 is encoded within the $\alpha 2$ exon⁹. Deletion of this region in the alignment only changes the homology from 33% to 31% while the significance score decreases from 6.98 to 4.75. *c*, Predicted structure of the MAdCAM-1 polypeptide. The closest relationships of the three immunoglobulin domains to other members of the immunoglobulin supergene family are indicated above each loop along with the location of the serine/threonine-rich mucin-like domain. A double disulphide bond is indicated in the first immunoglobulin domain to illustrate the unusual paired cysteines at each end, a feature that is shared with ICAM-1 and VCAM-1. The actual nature of cysteine bonding in these loops is unknown.



phocyte counter-receptors LFA1/ α L β 2 (ref. 18) and VLA4/ $\alpha 4 \beta 1$ (ref. 19), raising the possibility that an integrin-binding motif may also lie within this domain of MAdCAM-1. The most likely candidate integrin receptor for MAdCAM-1 is $\alpha 4 \beta 7$, which has been implicated in lymphocyte binding to PP-HEV in frozen sections²⁰. The second immunoglobulin domain is most closely related to the fifth domain of mouse VCAM-1²¹, which was surprising because the fifth domain of MAdCAM-1 is a V-type domain whereas the fifth domain of VCAM-1 is a C2-type domain^{15,16}. Nonetheless, the homology may be significant given 30% residue identity and the fact that this VCAM-1 domain is the only immunoglobulin superfamily member that shows similarity in the homology search. The third immunoglobulin domain of the MAdCAM-1 displays an intriguing homology to the $\alpha 2$ constant region immunoglobulin loop of human and gorilla IgA1^{8,9} (Fig. 3*b, c*). This homology is particularly interesting as it relates two molecules that are highly expressed in mucosal tissues, and which both interact with T cells involved in mucosal immunity. $\alpha 2$ mediates IgA interactions with the poly-immunoglobulin Fc receptor²², and probably with IgA-specific Fc α receptors expressed by mucosal T-cell subsets; it is interesting to speculate that the homologous MAdCAM-1 domain may also be able to interact with the Fc α receptor or related lymphocyte surface receptors. Such an interaction could strengthen and provide additional specificity to adhesion by Fc α receptor-positive mucosal T cells.

In addition to the immunoglobulin domains, the MAdCAM-1 sequence predicts a 37 amino-acid region that lies between domains two and three which is extremely rich in its serine/threonine content (41%, Figs 1 and 3*c*), exhibiting a sequence characteristic of mucins²³. This domain contains numerous potential sites for addition of O-linked carbohydrates which are known to modify MAdCAM-1¹² and may have functional significance in lymphocyte-HEV interactions. Recently, a 50K serine/threonine-rich mucin-like glycoprotein, GlyCAM-

1, has been described which is highly expressed by and secreted from lymph node HEV²⁴. GlyCAM-1 displays O-linked carbohydrates characteristic of the peripheral lymph node addressin (PNAd). These carbohydrates are recognized by the L-selectin lymph node homing receptor, and by anti-PNAd monoclonal antibody MECA-79 (ref. 25). PNAd-associated epitopes recognized by MECA-79 can also decorate MAdCAM-1 isolated from mesenteric lymph nodes, whose HEV express both PNAd and MAdCAM-1 at high levels (E. Berg *et al.*, manuscript submitted). Thus the mucin-like MAdCAM-1 domain may be a site for addition of selectin-binding carbohydrates important for recruitment of lymphocyte subsets to the mesenteric lymph node and possibly, at a lower level, to Peyer's patches as well.

We conclude that the tandem domain structure of the mucosal addressin may be able to support multiple molecular events involved in the complex process of lymphocyte-HEV recognition and adhesion *in vivo*. □

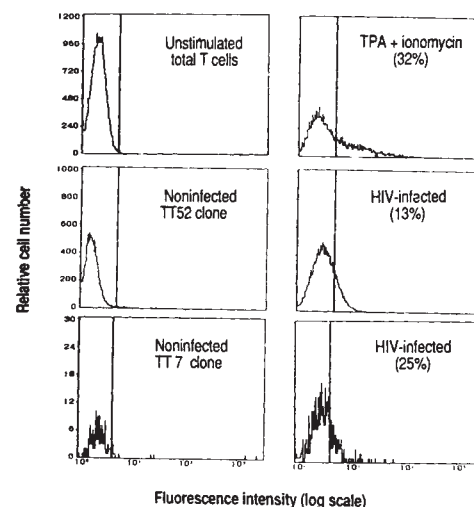
Received 30 October 1992; accepted 15 March 1993.

- Butcher, E. C., Scollay, R. G. & Weissman, I. L. *Eur. J. Immun.* **10**, 551-561 (1980).
- Butcher, E. C. *Curr. Topics Microbiol. Immun.* **133**, 2961-2965 (1986).
- Pickler, L. J. & Butcher, E. C. *Ann. Rev. Immun.* **10**, 561-591 (1992).
- Streeter, P. R., Berg, E. L., Rouse, B. N., Bargatze, R. F. & Butcher, E. C. *Nature* **331**, 41-46 (1988).
- Nakache, M., Berg, E. L., Streeter, P. R. & Butcher, E. C. *Nature* **337**, 179-181 (1989).
- Kita, Y. *et al. Biochim. biophys. Acta* **1131**, 108-110 (1992).
- Osborn, L. *et al. Cell* **59**, 1203-1211 (1989).
- Kawamura, S., Omoto, K. & Ueda, S. *Nucleic Acids Res.* **17**, 6732 (1989).
- Flanagan, J. G., Lefranc, M. P. & Rabbitts, T. H. *Cell* **36**, 681-688 (1984).
- Montesano, R. *et al. Cell* **62**, 435-445 (1990).
- von Heijne, G. *Nucleic Acids Res.* **14**, 4683-4690 (1986).
- Berg, E. L., Pickler, L. J., Robinson, M. K., Streeter, P. R. & Butcher, E. C. *Cellular and Molecular Mechanisms of Inflammation*, Vol. 2, 111-129 (Academic, San Diego, 1991).
- Seed, B. & Aruffo, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3365-3369 (1987).
- Simonet-Martin, M. T., Schaaf-Lafontaine, N. & Firket, H. *C. R. Soc. Biol.* **180**, 229-233 (1986).
- Williams, A. F. & Barclay, A. N. *Rev. Immun.* **6**, 381-405 (1988).
- Hunkapiller, T. & Hood, L. *Adv. Immun.* **44**, 1-63 (1989).
- Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435-1441 (1985).
- Staunton, D. E., Dustin, M. L., Erickson, H. P. & Springer, T. A. *Cell* **61**, 243-254 (1990).
- Osborn, L., Viskiallo, C. & Benjamin, C. D. *J. exp. Med.* **176**, 99-107 (1992).

20. Holtzman, B., McIntyre, B. W. & Weissman, I. L. *Cell* **56**, 37–46 (1989).
 21. Hession, C. et al. *Biochem. biophys. Res. Commun.* **183**, 163–169 (1992).
 22. Frutiger, S., Hughes, G. J., Hanley, W. C., Kingzette, M. & Jaton, J. *J. biol. Chem.* **261**, 6673–6681 (1986).
 23. Jentoft, N. *Trends Biochem. Sci.* **15**, 291–294 (1990).
 24. Lasky, L. A. et al. *Cell* **69**, 927–938 (1992).
 25. Berg, E. L., Robinson, M. K., Warnock, R. A. & Butcher, E. C. *J. Cell Biol.* **114**, 343–349 (1991).
 26. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory Press, New York, 1990).
 27. Gallatin, W. M., Weissman, I. L. & Butcher, E. C. *Nature* **303**, 30–34 (1983).
 28. Jalkanen, S. T., Bargatzke, R. F., de los Toyos, J. & Butcher, E. C. *J. Cell Biol.* **105**, 983–990 (1987).

ACKNOWLEDGEMENTS. We thank W. Rissau for the gift of bEnd.3 cells, L. Sikorski for identifying the expression of MAdCAM-1 on bEnd.3 cells, E. Berg for production of polyclonal anti-addressin sera, J. Bellenson at the protein and nucleic acid facility of the Center for Molecular and Genetic Medicine for DNA sequencing and T. Klinger for assistance with homology searches. Financial support was provided by the NIH, the Veterans Administration, and the Smith Kline Beecham research initiative.

T cell clone	TNF- α RNA expression		TNF- α secretion (U ml ⁻¹)	
	HIV ⁻	HIV ⁺	HIV ⁻	HIV ⁺
TT52			<50 (874)	518 (2,640)
TT 7			<50 (1,066)	638 (7,638)
TT8			<50 (130)	250 (1,840)



Membrane tumour necrosis factor- α is involved in the polyclonal B-cell activation induced by HIV-infected human T cells

Donatella Macchia, Fabio Almerigogna*, Paola Parronchi, Adriana Ravina, Enrico Maggi & Sergio Romagnani†

Division of Clinical Immunology and Allergology, University of Florence, Viale le Morgagni 85, 50134 Florence, Italy

* Istituto di Clinica Medica I, University of Pisa, Policlinico di Santa Chiara, 56100 Pisa, Italy

INFECTION of CD4⁺ T cells by human immune deficiency virus-1 (HIV-1) causes severe dysfunction of cellular immunity^{1–3}, but paradoxically results in intense polyclonal activation of B cells, possibly accounting for both hypergammaglobulinaemia and frequent development of B-cell malignancies seen in HIV-infected patients^{4–7}. We have reported that human CD4⁺ T-cell clones infected with HIV *in vitro* markedly stimulate immunoglobulin synthesis by B cells through a non-cognate, contact-dependent mechanism⁸. We show here that HIV-infected T-cell clones do not express the CD40 ligand (CD40L), a molecule critical for non-cognate B-cell activation⁹, but a small proportion of them do express membrane tumour-necrosis factor (TNF)- α . The ability of HIV-infected T-cell clones to induce polyclonal B-cell activation appears to be restricted to TNF- α -positive T blasts and is inhibited by antibodies against both TNF- α and TNF- α receptor. Freshly isolated CD4⁺ T cells from HIV-infected individuals express TNF- α on the cell membrane and induce TNF- α -mediated immunoglobulin production by B cells. Thus, membrane TNF- α seems to be involved in the polyclonal B-cell activation induced by HIV-infected T cells.

The CD40L on the activated Th cell⁹ is an attractive candidate for relaying the critical non-cognate activation signal to the B cell. When assessed with a soluble CD40 fusion protein⁹, however, non-infected and HIV-infected T-cell clones were both found to express CD40L in response to stimulation with an anti-CD3 monoclonal antibody, whereas unstimulated non-infected and HIV-infected T-cell clones did not. Furthermore, addition in culture of the soluble CD40 fusion protein did not inhibit polyclonal immunoglobulin production induced by HIV-infected T-cell clones (data not shown). To find an alternative

FIG. 1 TNF- α secretion and expression of membrane TNF- α by HIV-infected T-cell clones. CD4⁺ T-cell clones specific for tetanus toxoid or recombinant Poa pratensis group IX (Poa p IX) allergen were generated from peripheral blood lymphocytes (PBL) of HIV-seronegative donors by limiting-dilution cloning as described^{8,16}. T blasts from each clone were then divided into two equal aliquots and co-cultured with irradiated (6,000r) HIV-infected or non-infected PBL, respectively. Co-culturing was done in the presence of phytohaemagglutinin (PHA) (1% v/v), interleukin-2 (20 u ml⁻¹) and polybrene (5 μ g ml⁻¹)⁸. On day 14, HIV infection of T-cell clones was assessed by either polymerase chain reaction (PCR) analysis of T-cell clones or measurement of both reverse transcriptase activity and p24 antigen in T-cell clone supernatants, as described⁸. Top, three tetanus toxoid-specific (TT52, TT7, TT8) HIV-infected T-cell clones (HIV⁺), as well as their non-infected counterparts (HIV⁻), were examined for TNF- α RNA expression by slot hybridization analysis¹⁶. Cytoplasmic RNA was prepared by cell lysis using Nonidet-P40. Pre-hybridization of the nylon membrane, preparation by 5'-end labelling of the ³²P-labelled human TNF- α oligonucleotide probe (specific activity 1–2 $\times$ 10⁸ c.p.m. ng⁻¹), hybridization and autoradiography have all been described¹⁶. TNF- α protein secreted in the cell-free supernatants of the same T-cell clones after 24 h culture in the absence of any stimulant or in the presence of 12-O-tetradecanoyl phorbol 13-acetate (TPA; 10 ng ml⁻¹) plus anti-CD3 antibody (100 ng ml⁻¹) was quantitated using a commercial ELISA (Quantikine; RD Systems, Minneapolis). Values in parentheses represent TNF- α levels found in supernatants from stimulated T-cell clones. Bottom, to assess TT52 and TT7 T-cell clones for expression of membrane TNF- α , clonal T blasts were incubated for 45 min at 4°C with anti-TNF- α antibody B154.2.1, washed three times and then incubated with FITC-conjugated anti-mouse IgG goat antibody (Becton-Dickinson). As a positive control, freshly prepared PB T cells from the same donor, stimulated with TPA (1 ng ml⁻¹) and the calcium ionophore ionomycin (0.5 μ M) as described¹⁰, were also tested for membrane TNF- α . T cells positive for membrane TNF- α were determined by FACS analysis using T cells incubated with purified mouse IgG to determine the background fluorescence signal. Cell treatment with low-pH buffer¹⁷ was used to exclude contribution of soluble TNF- α bound to its receptor. Data are from a representative two of seven HIV-infected unstimulated T-cell clones (4 specific for tetanus toxoid and 3 specific for Poa p IX).

† To whom correspondence should be addressed.

T-B cell-to-cell interaction pathway evoked by HIV infection, we investigated the role of membrane TNF- α , because (1) normal human T cells express surface TNF- α in response to synergistic combination of ionomycin and 12-*O*-tetradecanoyl phorbol-13-acetate, an activator of protein kinase C¹⁰; and (2) HIV-infected T-cell clones consistently express TNF- α RNA and secrete TNF- α in the absence of any stimulation, whereas their non-infected counterparts do not. We found that a variable proportion (5–25%) of blasts from HIV-infected T-cell clones express membrane TNF- α , whereas no TNF- α expression was ever observed on the membrane of their non-infected counterparts (Fig. 1). We therefore examined the effects of a goat polyclonal antibody specific to TNF- α and of a combination of two monoclonal antibodies specific for TNF- α receptors on B-cell proliferation and immunoglobulin production induced by HIV-infected T-cell clones. As shown in Table 1, these antibodies had no inhibitory effect on B-cell proliferation or on immunoglobulin production induced by anti-CD3-stimulated non-HIV-infected T-cell clones, but they strongly and consistently inhibited the B-cell activating property of HIV-infected T-cell clones.

To confirm that the polyclonal B-cell activation induced by HIV-infected T-cell clones was related to HIV-induced membrane TNF- α expression on a proportion of their T blasts, we fractionated three HIV-infected T-cell clones into TNF- α^+ and

TNF- α^- cells by sorting with anti-TNF- α monoclonal antibody. TNF- α^- , TNF- α^+ , and unsorted cells were then assessed for their ability to induce B-cell proliferation and immunoglobulin synthesis by autologous B cells. As shown in Fig. 2, only unsorted cells and TNF- α^+ , but not TNF- α^- cells induced both proliferation and immunoglobulin synthesis by autologous B cells, and the activity of TNF- α^+ cells was comparable to, or even higher than, that of unsorted cells, despite using anti-TNF- α antibody in the sorting procedure. This inability of anti-TNF- α antibody to inhibit polyclonal B-cell activation could be due to incomplete masking of all TNF- α determinants and/or to resynthesis of new membrane TNF- α molecules during culture.

To verify that what occurs during acute infection *in vitro* of CD4⁺ T-cell clones relates to the situation *in vivo*, we assessed the expression of membrane TNF- α by CD4⁺ T cell-enriched suspensions freshly isolated from the peripheral blood of 17 HIV-seropositive individuals and from 14 healthy HIV-seronegative controls (technique described in Table 2 legend). There was a statistically significant ($P < 0.05$) difference in the mean values ($\pm$ s.d.) of CD4⁺ T cells expressing membrane TNF- α between HIV-seropositive individuals (2.4 ± 3.9) and controls (0.1 ± 0.3). Interestingly, the proportion of TNF- α^+ in HIV-seropositive individuals was inversely related to the number of peripheral blood CD4⁺ T cells ($r = -0.53$); the two HIV-seropositive subjects with the most TNF- α^+ cells had full-blown

TABLE 1 Inhibitory effect of antibodies against TNF- α and its receptor on B-cell proliferation and immunoglobulin production induced by HIV-infected T-cell clones

Clonal T blasts added to B cells	³ H-thymidine uptake (c.p.m. $\times 10^{-3}$)	Immunoglobulin synthesis (μ g ml ⁻¹)	
		IgM	IgA
None	<0.5	<0.1	<0.5
Non-infected unstimulated	<0.5	<0.1	<0.5
Non-infected anti-CD3-stimulated	13 $\pm$ 3	9 $\pm$ 1	11 $\pm$ 3
Non-infected anti-CD3-stimulated + anti-TNF- α	12 $\pm$ 2	10 $\pm$ 2	9 $\pm$ 2
Non-infected anti-CD3-stimulated + anti-TNF- α -receptor	14 $\pm$ 3	10 $\pm$ 2	12 $\pm$ 3
HIV-infected unstimulated	25 $\pm$ 3	18 $\pm$ 3	31 $\pm$ 4
HIV-infected unstimulated + anti-TNF- α	3 $\pm$ 1	2 $\pm$ 1	4 $\pm$ 2
HIV-infected unstimulated + anti-TNF- α -receptor	<0.5	1 $\pm$ 1	2 $\pm$ 1

B-cell-enriched suspensions were prepared from the peripheral blood of the same subjects as were used as donors of T-cell clones⁸. For the proliferation assay, unstimulated HIV-infected T-cell clones, as well as their non-infected unstimulated or anti-CD3-stimulated (100 ng ml⁻¹) counterparts, were irradiated (6,000r) and co-cultured for 5 days at a concentration of 2×10^4 cells per ml with autologous B cells (1×10^5 per ml). 16 h before collection, cultures were pulsed with 0.5 μ Ci ³H-thymidine (Amersham) and the counts incorporated measured by liquid scintillation counting. For immunoglobulin production assay, HIV-infected T-cell clones, as well as their unstimulated or anti-CD3-stimulated (100 ng ml⁻¹) non-infected counterparts, were co-cultured for 8 d at 2×10^4 cells per ml with autologous B cells (1×10^5 per ml). Cell-free culture supernatants were assayed for IgM and IgA by radioimmunoassay⁸. The goat polyclonal anti-TNF- α antibody (Janssen, Belgium) was added in culture at 1 μ g ml⁻¹. The two anti-TNF- α -receptor monoclonal antibodies Utr-1 and Htr-9, reactive against type A and type B TNF- α receptors respectively¹⁵, were used in combination (v/v) at a final concentration of 1 μ g ml⁻¹. The results represent mean values ($\pm$ s.e.) from six different clones (three specific for tetanus toxoid, and three specific for Poa p IX) derived from two different donors. Non-immune goat serum and mouse IgG, added at the same concentrations, did not inhibit either the B-cell proliferative response or immunoglobulin production induced by HIV-infected T-cell clones.

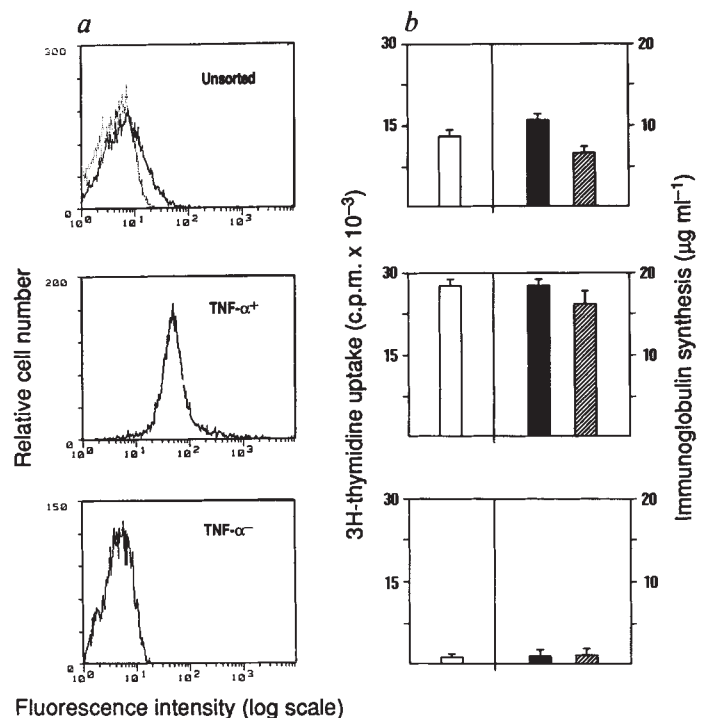


FIG. 2 Effect of sorting with anti-TNF- α antibody on the ability of HIV-infected T-cell clones to induce polyclonal B-cell activation. *a*, T blasts from a Poa p IX-specific HIV-infected T-cell clone were reacted with anti-TNF- α antibody (B154.2.1), followed by FITC-labelled anti-mouse IgG goat Ab (Becton-Dickinson). Sorting was done on a Becton-Dickinson FACStar. Reanalysis showed that FITC-labelled cells were 6% in the starting population (unsorted), less than 0.1% in the TNF- α^- and 93% in the TNF- α^+ cell fraction. Dotted line represents the FACS profile obtained by incubating unsorted T blasts with normal mouse IgG, followed by FITC-labelled anti-mouse IgG goat antibody. *b*, Unsorted and sorted T blasts were co-cultured with autologous B cells and assessed for their ability to induce B-cell proliferation (white columns), IgM (black columns) and IgA synthesis (shaded columns), as described in Table 1 legend. Results represent mean values ($\pm$ s.d.) of triplicate determinations. Comparable results were obtained in two other sorting experiments.

TABLE 2 Immunoglobulin induced in autologous B cells by CD4⁺ T cells from lymph node biopsy of an HIV-seropositive individual

Co-culture with B cells	Immunoglobulin synthesis ($\mu\text{g ml}^{-1}$)			CD4 ⁺ T cells with membrane TNF- α (%)
	IgM	IgG	IgA	
Medium	0.6	0.7	0.4	
PB CD4 ⁺ T cells	1.2	0.3	0.7	<0.1
LN CD4 ⁺ T cells	5.6	2.6	3.9	2.5
+ anti-TNF- α - receptor	0.9	0.2	0.5	

A biopsy specimen was taken for diagnostic purposes from the axillary lymph nodes (LN) of a HIV-seropositive subject having normal numbers of peripheral blood (PB) CD4⁺ T cells ($860 \text{ cells mm}^{-3}$). Histological examination revealed a pattern of nonspecific adenitis with strong follicular hyperplasia, suggesting classical HIV-related lymphadenopathy. The specimen was teased and the resulting cell suspension centrifuged on a Ficoll-Hypaque density gradient. CD4⁺ T cells were purified from LN and PB derived from the same subject by incubating mononuclear cell suspensions with a mixture of monoclonal antibodies reactive with monocytes (anti-CD11b, OKM1; Ortho Pharmaceutical), large granular lymphocytes (anti-CD16, Leu 19; Becton-Dickinson), B lymphocytes (anti-CD20, B1; Kontron Instruments) and CD8⁺ cells (anti-CD8; OKT8, Ortho Pharmaceutical), followed by incubation with immunomagnetic beads (Dynabeads; Dynal, Oslo) covalently bound to goat anti-mouse IgG antibody and separation with a magnetic cell separator (Dynal). With this technique, cell suspensions recovered are usually >95% CD4⁺ T cells. Preparation of B-cell-enriched suspensions from the PB of the same subject and immunoglobulin production assays are reported in the legend to Table 1. The two anti-TNF- α -receptor antibody Utr-1 and Htr-9 were used in combination (Table 1 legend). To quantify the CD4⁺ T cells expressing membrane TNF- α , cells were first incubated with anti-TNF- α antibody and then with FITC-conjugated anti-mouse IgG goat antibody (see also Fig. 1 legend), followed by fixation with 0.5% paraformaldehyde and staining with phycoerythrin-conjugated anti-CD4 antibody (Leu 3a; Becton-Dickinson).

disease (data not shown). These results may be consistent with the observation that in HIV-infected individuals viral replication is minimal at blood level for many years after initial infection, but increases in the more advanced stages¹¹. Recently it has been shown that massive HIV replication occurs in lymph nodes even early after initial infection^{12,13}. As polyclonal B-cell activation may appear early after infection¹¹, HIV-infected CD4⁺ T cells in lymph node follicles and interfollicular areas may be responsible for activation of B cells present in these micro-environments. We therefore tested the ability of highly purified CD4⁺ T cells, isolated from a lymph node biopsy specimen of an HIV-seropositive patient having virtually normal numbers ($860 \text{ cells per mm}^3$) of peripheral blood CD4⁺ T cells, to induce immunoglobulin synthesis in autologous B cells. As shown in Table 2, lymph node but not peripheral blood CD4⁺ T cells from this subject were able to induce synthesis of detectable amounts of IgM, IgG and IgA by autologous B cells; this effect was strongly inhibited in the presence of anti-TNF- α -receptor antibody.

Taken together, our results strongly support the possibility that the polyclonal B cell activation seen in HIV-infected subjects is mediated through the non-antigen-specific, MHC-unrestricted, membrane TNF- α -mediated pathway described here. Whether membrane TNF- α alone is sufficient to activate the B cells or whether it acts together with some HIV-related signal, remains to be established. But activation of this abnormal pathway, apparently distinct from that operating in anti-CD3 (or antigen)-activated normal T cells, which probably relies on CD40L expression⁹, may explain at least in part the hypergammaglobulinaemia and other features related to polyclonal B-cell activation⁴⁻⁷. Moreover, continuous B-cell stimulation by HIV-infected T cells by the membrane TNF- α pathway might

also play a part in the development of B-cell malignancies¹⁴ in HIV-infected patients. □

Received 23 October 1992; accepted 27 April 1993.

1. Lane, H. C. *et al.* *New Engl. J. Med.* **313**, 79-84 (1985).
2. Fahey, J. L. *et al.* *Am. J. Med.* **76**, 95-100 (1984).
3. Bowen, D. L., Lane, H. C. & Fauci, A. S. *Ann. Intern. Med.* **103**, 704-709 (1985).
4. Amman, A. J. *et al.* *Clin. Immun. Immunopath.* **27**, 315-325 (1983).
5. Euler, H. H., Kern, P., Löffler, H. & Dietrich, M. *Clin. exp. Immun.* **59**, 267-275 (1985).
6. Kopelman, R. G. & Zolla-Pazner, S. *Am. J. Med.* **84**, 82-88 (1988).
7. Lane, H. C. *et al.* *New Engl. J. Med.* **309**, 453-458 (1983).
8. Macchia, D. *et al.* *J. Immun.* **146**, 3413-3418 (1991).
9. Armitage, R. J. *et al.* *Nature* **357**, 80-82 (1992).
10. Kinkhaowala, M. *et al.* *J. exp. Med.* **171**, 941-946 (1990).
11. Terpstra, F. G. *et al.* *Eur. J. Immun.* **19**, 667-673 (1989).
12. Pantaleo, G. *et al.* *Nature* **362**, 355-358 (1993).
13. Embretson, J. *et al.* *Nature* **362**, 359-362 (1993).
14. Amadori, A. & Chieco Bianchi, G. *Immun. Today* **11**, 374-379 (1990).
15. Shalaby, M. R. *et al.* *J. exp. Med.* **172**, 1517-1520 (1990).
16. Parronchi, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 4538-4593 (1991).
17. Kulczycki, A. & Metzger, H. *J. exp. Med.* **140**, 1676-1685 (1974).

ACKNOWLEDGEMENTS. We thank S. Mohapatra for recombinant Poa p IX allergen; the Istituto Sieroterapico e Vaccinogeno Sclavo, Siena, for tetanus toxoid; S. Gillis for CD40-Ig; G. Trinchieri for anti-TNF- α mAb B154.2.1; and M. Brockhaus for anti-TNF- α receptor mAbs Utr-1 and Htr-9. This work was supported by grants from the AIRC, Istituto Superiore di Sanita and CNR.

The origin of HIV-1 isolate HTLV-IIIB

Sheng-Yung P. Chang, Barbara H. Bowman,
Judith B. Weiss, Rebeca E. Garcia & Thomas J. White*

Roche Molecular Systems Inc., 1145 Atlantic Avenue, Alameda, California 94501, USA

THE striking similarity between the first two human immunodeficiency virus type 1 (HIV-1) isolates Lai/LAV (formerly LAV, isolated at the Pasteur Institute^{1,2}) and Lai/IIIB (formerly HTLV-IIIB, reported to be isolated from a pooled culture at the Laboratory of Tumor Cell Biology (LTCB) of the National Cancer Institute^{3,4}) provoked considerable controversy in light of the high level of variability found among subsequent HIV-1 isolates⁵. In November 1990, the Office of Scientific Integrity at the National Institutes of Health commissioned our group to analyse archival samples established at the Pasteur Institute and LTCB between 1983 and 1985. Retrospective analyses^{6,7} have shown that contamination of a culture derived from patient BRU by one from patient LAI was responsible for the provenance of HIV-1 Lai/LAV; the contaminated culture (M2T-/B) was sent to LTCB in September 1983^{6,7}. Our goals were to determine which HIV-1 variants were present in the samples and the sequence diversity among HIV-1 isolates from the earliest stages of the AIDS epidemic. We examined archival specimens and report here the detection of six novel HIV-1 sequences in the cultures used to establish the pool: none is closely related to HIV-1 Lai/IIIB. A sample derived from patient LAI contained variants of both HIV-1 Lai/IIIB and HIV-1 Lai/LAV, and a sequence identical to a variant of HIV-1 Lai/IIIB was detected in the contaminated M2T-/B culture. We conclude that the pool, and probably another LTCB culture, MoV, were contaminated between October 1983 and early 1984 by variants of HIV-1 Lai from the M2T-/B culture. Therefore, the origin of the HIV-1 Lai/IIIB isolate also was patient LAI.

Our study analyses the variability and relationships among the HIV-1 nucleotide sequences in 37 coded archival samples associated with the isolation of HIV-1 Lai/LAV and HIV-1 Lai/IIIB. To avoid starting from the assumption that variants of HIV-1 Lai/IIIB and HIV-1 Lai/LAV were true quasiespecies (variants from a single isolate)⁸, we chose to analyse the hyper-variable regions V1 and V2 (V1/V2) in the *env* gene that distin-

* To whom correspondence should be addressed.

TABLE 1 Results of PCR assays

Code	Origin	Presence of HIV-1 by <i>gag</i> PCR	Sequence of <i>env</i> V1/V2 region	Number of clones sequenced	Code	Origin	Presence of HIV-1 by <i>gag</i> PCR	Sequence of <i>env</i> V1/V2 region	Number of clones sequenced
Group I. Derived from patient BRU					Group III. Related to the pool established at LTCB				
A1	JBB/LAV (DNA)	+	HIV-1 Bru	9	A8	Pool-related (SE)	+	Novel	7
A2	Pasteur Institute	+	HIV-1 Bru	19	A9	Pool-related (CH)	+	—	10
A3	JBB/LAV	+	HIV-1 Bru	19	A10	Pool-related (MA)	+	Novel	10
A4	passaged at LTCB	—	—	—	A11	Pool-related (TW)	—	—	—
A5	JBB/LAV	+	HIV-1 Bru	20	A12	Pool-related (DB)	—	—	—
B7	Pasteur Institute	+	HIV-1 Bru	10	A13	Pool-related (EC)	—	—	—
C1	JBB/LAV	+	HIV-1 Bru	9	A14	Pool-related (WA)	—	—	—
	Pasteur Institute	—	—	—	A15	Pool-related (RW)	+	Novel	8
		—	—	—	A16	Pool-related (GM)	+	Novel	9
		—	—	—	A17	Pool-related (RAW)	+	Novel	7
		—	—	—	B5	Donor (CH) of A9	+	Novel	14
		—	—	—	B6	Donor (EC) of A13	—	—	—
Group II. Established at Pasteur Institute					Group IV. Related to the isolation of HIV-1 Lai/IIIB (HTLV-IIIB)				
A6	Traceable to	+	HIV-1 Lai/LAV	14	A18	H17/HTLV-IIIB	—	—	—
	M2T-/B passaged	—	HIV-1 Lai/IIIB	6	A19	H9/HTLV-IIIB	+	HIV-1 Lai/IIIB	10
	by M. Martin	—	—	—	A20	H9/HTLV-IIIB	+	HIV-1 Lai/IIIB	10
A7	Traceable to	+	HIV-1 Lai/LAV	7	A21	H9/HTLV-IIIB	+	HIV-1 Lai/IIIB	8
	M2T-/B passaged	—	HIV-1 Lai/IIIB	12	A21A	H9/HTLV-IIIB	+	HIV-1 Lai/IIIB	20
	by M. Martin	—	—	—	Group V. Related to other cultures established at LTCB				
B1	Traceable to M2T-/B	+	HIV-1 Lai/LAV	11	A22	H4/MoV	+	HIV-1 Lai/IIIB	10
	sent to CDC	—	HIV-1 Lai/IIIB	4	A23	A putative donor	—	—	—
B2	Traceable to M2T-/B	+	HIV-1 Lai/LAV	3		(HM) of MoV	—	—	—
	sent to CDC	—	HIV-1 Lai/IIIB	15	A24	HT/RF	+	HIV-1 RF	10
B3	Traceable to M2T-/B	+	HIV-1 Lai/LAV	4			—	—	—
	passaged at CDC	—	HIV-1 Lai/IIIB	16			—	—	—
B4	Traceable to M2T-/B	+	HIV-1 Lai/LAV	13			—	—	—
	passaged at CDC	—	HIV-1 Lai/IIIB	7			—	—	—
B8	M2T-/B	+	HIV-1 Bru	3			—	—	—
		—	HIV-1 Lai/LAV	5			—	—	—
		—	HIV-1 Lai/IIIB	4			—	—	—
B9	B-LAV	+	HIV-1 Lai/IIIB	10			—	—	—
C2	JBB4/IDAV-2 (DL)	+	Novel	9			—	—	—
C3	JBB4/IDAV-1 (LAI)	+	HIV-1 Lai/LAV	18			—	—	—
		—	HIV-1 Lai/IIIB	10			—	—	—

Coded samples and their origins were provided by the Office of Scientific Integrity (OSI) of the National Institutes of Health (NIH). A highly sensitive oligomer hybridization PCR assay, which incorporates dUTP and uracil-*N*-glycosylase in the amplification to avoid false-positive results, was used to detect the relatively conserved *gag* region^{21,22}. A nested-PCR system was used to amplify the *env* V1/V2 region (corresponding to positions 6,583–6,851 in HXB2²³). Sequences with the signature (two 3-nucleotide deletions and several substitutions) of the reported JBB/LAV⁷ were classified as HIV-1 Bru. Sequences with the 15-nucleotide deletion (positions 6,653–6,667 in HXB2) and 5 of the 6 signature substitutions associated with variants of HIV-1 Lai/IIIB^{4,9,10} were classified as HIV-1 Lai/IIIB. Sequences with a 15-nucleotide duplication as observed in J19 (LAV-1a)² were classified as HIV-1 Lai/LAV. Samples B5 and B6 were sera from donors of samples A9 and A13, respectively. The probability of missing a specific variant present in a population of clones can be closely approximated by $P = (1 - p)^n$, where p is the fraction of the specific variant in the population and n is the number of clones randomly selected and analysed out of a large population. The distribution of types of variants in two different small samples from the same population would not necessarily be expected to be the same or to reflect the distribution of variants in the original population. Sequences of eight unique clones derived from sample B8, accession numbers M96463 to M96470 were deposited in GenBank. Sequences of 22 unique clones derived from sample C3, accession numbers M96471 to 96492, were deposited in GenBank. Nucleic acids were extracted from samples and analysed by PCR in pairs (three at a time in two cases) to minimize sample-to-sample contamination. Precautions were taken to minimize contamination during the preparation of samples for PCR, the synthesis of complementary DNA, and the amplification of template DNA²³. Template DNA from cell culture samples A8–A16, A18–A22, A21A and A24, was extracted with minor modifications of the procedure described in ref. 24. The cell lysates of samples A9, A11–A14, A18, and A21A did not yield any PCR product, therefore the supernatant fluid fractions from these samples were included with the group of cell-free supernatant fluid samples. The total nucleic acids in the plasma sample (A23), serum samples (A4, A17, B5 and B6) and cell-free supernatant fluid samples (A2, A3, A5–A7, B1–B4 and A21A) were extracted with an 'IsoQuick' kit (MicroProbe Corporation (Bothell, WA)) using the procedures provided for the extraction of total nucleic acids. Serum or plasma or cell-free supernatant fluid (20–50 μ l) of the sample was used. Both the reverse transcriptase-treated sample²⁵ (proviral DNA/cDNA) and an equivalent amount of untreated sample (proviral DNA only) were used in the PCR. No distinct difference in sequence variability was observed between the sequences derived from amplification of reverse transcriptase-treated and untreated sample. The supernatant fluid fraction of six samples (B7–B9, C1, C2–C3) was used directly as the source of template after the heat treatment⁶. The *gag* PCR assay: primer SK145 and primer SK431 (with an additional T at the 3' end of SK150) with dUTP substituted for TTP were used for PCR. The presence of the 142-bp amplified product was detected by probe SK102^{21,22}. Ten molecules of positive control linearized plasmid DNA pSYC1857 (containing rearranged, inactivated HIV26 sequences)²⁵ were consistently detected in this assay. The signal of the *gag* PCR assay of sample A9 was much weaker than that of obtained from 10 molecules of pSYC1857. The fidelity of *Taq* DNA polymerase under the conditions for the nested *env* PCR was determined²⁵. The average error frequency of *Taq* DNA polymerase after two rounds of 35 cycles of amplification was 1 in 720 using 0.2 mM of each dNTP. This error frequency corresponds to a misincorporation rate of 4.0×10^{-5} per cycle (assuming 70 doublings in two rounds of amplification). The mutations observed were transitions. The nested *env* PCR assay used external primers SK452 (5'-GCGGAATTCGATATAATCAGTTTATGGGAT-3') and SK455 (5'-ATACTGCAGGCCTAATTCATGTGTACATT-3') and internal primers SK453 (5'-TGGGATCCAAG-CCTAAAGCCATGTGT-3') and SK454 (5'-GCGAAGCTTAATGTATGGGAATTGGCTCAA-3'). A high-temperature 'activation' step²⁶ was included to reduce nonspecific annealing and extension of the primers. The conditions for PCR were described in ref. 25. The number of cycles for the second amplification was reduced to 20 for samples B7 and B8, and to 15 for samples B1 to B4 to prevent PCR products from reaching plateau. A positive control of 10 molecules of pSYC1857 DNA, and a negative control without DNA were always included. The agarose gel-fractionated nested-PCR products were precipitated, digested with restriction enzymes, and cloned into the *Hind*III and *Bam*HI sites of pUC18 plasmid DNA²⁵. The size(s) of inserts in the transformants were confirmed by PCR using primers that flanked the polylinker region²⁵. The nucleotide sequences of the inserts were determined by the dideoxy chain-termination method. The nucleotide sequence of a total of 380 clones was determined (~125 kb).

guish HIV-1 Lai/IIIB from HIV-1 Lai/LAV. A 15-nucleotide deletion and 6 single-nucleotide substitutions are shared by all variants of HIV-1 Lai/IIIB (refs 4, 9, 10). But we note that the variant NL43 (refs 5, 11) of HIV-1 Lai/LAV also has the signature sequence of HIV-1 Lai/IIIB in this region. The results of the polymerase chain reaction (PCR)¹² analyses are divided into five groups and summarized in Table 1.

The first group consists of a serum sample (A4) from patient BRU, who was originally thought to be the source of HIV-1 Lai/LAV, and six samples derived from the JBB/LAV culture (from patient BRU)⁶. In agreement with previous analyses^{6,7}, only variants of HIV-1 Bru were detected in these JBB/LAV-derived samples. The absence of template for PCR in the serum sample was probably due to the instability of DNA/RNA during long-term storage and multiple freeze-thaw cycles.

The second group consists of cultures used at the Pasteur Institute during the period when HIV-1 Lai/LAV contaminated the culture derived from patient BRU⁶. Variants of both HIV-1 Lai/IIIB and HIV-1 Lai/LAV were detected in JBB/IDAV-1 (ref. 6; C3, from patient LAI). We detected not only variants of HIV-1 Bru and HIV-1 Lai/LAV, but also a sequence identical to that of HIV-1 Lai/IIIB variants BH10 and PV22 in a sample of the contaminated M2T-/B culture (B8) that had been frozen and stored at the Pasteur Institute since August 1983.

Seven cultures traceable to M2T-/B were also analysed. Variants of both HIV-1 Lai/IIIB and HIV-1 Lai/LAV were detected in six of these cultures. The absence of HIV-1 Bru could be the result of HIV-1 Lai/LAV and HIV-1 Lai/IIIB outgrowing HIV-1 Bru. Only variants of HIV-1 Lai/IIIB were detected in the seventh sample, a B-cell-adapted LAV culture⁶ (B9). Variants of HIV-1 Lai/LAV might have been selectively eliminated during the passage of this B-cell line. In another culture, JBB4/IDAV-2 (ref. 6; C2, from patient DL), only variants of a novel HIV-1 (DL) were detected.

A third group of samples addresses the efforts at LTCB that resulted in the isolation of HIV-1 Lai/IIIB³. The chronology of events that related to the establishment of the pool derived from 10 donor samples (10 cocultivation cultures) is summarized in Table 2. The *env* V1/V2 sequences in samples A8, A9/B5, A10, A15, A16 and A17 were novel and unique, that is, distinct from variants of HIV-1 Lai/IIIB, HIV-1 Lai/LAV, HIV-1 Bru and any HIV-1 sequences in the Los Alamos database⁵. The absence of HIV-1 sequence in samples A11, A12, A13/B6 and A14 is consistent with the medical history of the four donors. Retrospectively, all four patients were seronegative at the time when samples were taken. Three patients never had AIDS, and the donor of sample A12 seroconverted between June 1984 and June 1985. General degradation of DNA template or the presence of inhibitors of PCR in the samples was ruled out by the detection of the cellular HLA-DQ alpha gene (HLA-DQ A1)¹³ by PCR. Nine samples that went into the pool had eight different HLA-DQ A1 genotypes and the genotypes of two serum samples matched those of their respective cocultivation cultures.

If aliquots of any cultures (A8–A10) frozen before the arrival of M2T-/B at LTCB and contributed to the pool were found to contain variants of HIV-1 Lai/IIIB, it would have been the only strong evidence supporting the independent isolation of HIV-1 Lai/IIIB. But none of these three samples was observed to contain variants of HIV-1 Lai/IIIB.

The fourth group consists of cultures expected to contain HIV-1 Lai/IIIB, including a blind positive control (A19). Variants of HIV-1 Lai/IIIB were detected in these samples with the exception of sample A18. The general integrity of DNA and lack of inhibitors of PCR in sample A18 were confirmed by the detection of the HLA-DQ A1 gene. The presence of HIV-1 in H17 T cells such as sample A18 was described in refs 3 and 14.

The fifth group consists of the RF and MoV cultures established at LTCB during the same time period as the pool, and a plasma sample from a putative donor of the MoV culture (A23). Only variants of HIV-1 RF⁵ were detected in a second blind-

TABLE 2 Chronological events related to the isolation of HIV-1 Lai/IIIB

Related to the pool	
Date	
23/9/83	The supernatant fluid of the contaminated M2T-/B culture from the Pasteur Institute was received by LTCB ⁶ .
Late 10/83–1/84	M2T-/B supernatant fluid was used to infect various types of cells ⁷ .
15/11/83	The pool was started at LTCB by pooling three cocultivation cultures (samples A8, A9 and A10) derived from peripheral blood lymphocytes (PBLs) of three donors.
22/11/83	The same three cocultivation cultures (samples A8, A9 and A10) were added to the pool again.
16/12/83	Serum (sample B6) was collected from the donor of cocultivation culture sample A13.
2/1/84	Cocultivation cultures (samples A11–A16) derived from PBLs of seven additional donors were added to the pool. (Serum (sample A17) of the seventh donor was analysed in place of the unavailable cocultivation culture.)
2/84	HIV-1 Lai/IIIB (HTLV-IIIB) was isolated from the pool. T-cell clone H17 infected with HIV-1 Lai/IIIB was frozen (sample A18)*.
4/84	T-cell clone H9 infected with HIV-1 Lai/IIIB was sent to Frederick Cancer Research (samples A20).
5/84	T-cell clone H9 infected with HIV-1 Lai/IIIB was sent to Frederick Cancer Research (sample A21). Second passage of HIV-1 Lai/IIIB was sent to Biotech Research Laboratory (sample A21A).
3/1/85	Serum (sample B5) was collected from donor of the cocultivation culture sample A9.
1989	T-cell clone H9 was infected with HIV-1 Lai/IIIB (sample A19).
Related to other cultures established during the same period as the pool	
9/83	Plasma (sample A23) from a putative donor of MoV isolate was received by LTCB.
9/11/83	RF culture was started by cocultivating HT cells with PBLs from donor RF.
29/11/83	T-cell clone H4 was infected with the HUT78/MoV isolate.
2/84	The cocultivation RF culture was frozen (sample A24).
3/84	T-cell clone H4 infected with the MoV isolate was frozen (sample A22).

The history of all samples established at LTCB was provided by OSI. No sample of the pool is known to exist.

* The possibility that sample A18 was an aliquot of uninfected H17 T cells cannot be ruled out.

positive control, the RF culture (A24). Only variants of HIV-1 Lai/IIIB were detected in the MoV culture (A22). The absence of template for PCR in the sample from a putative donor might be due to the instability of DNA/RNA in the plasma. The possibility of contamination or replacement of the MoV culture with HIV-1 from M2T-/B or an HIV-1 Lai/IIIB-containing culture cannot be ruled out.

An evolutionary analysis was done to determine the relationships among the HIV-1 sequences described above. The evolutionary distance tree in Fig. 1 illustrates that the six novel sequences from cultures that formed the pool are quite different from each other and from variants of HIV-1 Lai/LAV, HIV-1 Lai/IIIB, HIV-1 Bru, and reference sequences of North American, Caribbean and African isolates⁵. One might have expected the degree of variability among HIV-1 isolates collected early in the AIDS epidemic to be restricted. But the variability in the *env* V1/V2 region was already considerable as early as 1983–1984, and was quite sufficient to distinguish HIV-1 viruses from different individuals.

The very close phylogenetic relationship of HIV-1 Lai/LAV and HIV-1 Lai/IIIB is consistent with their having been derived

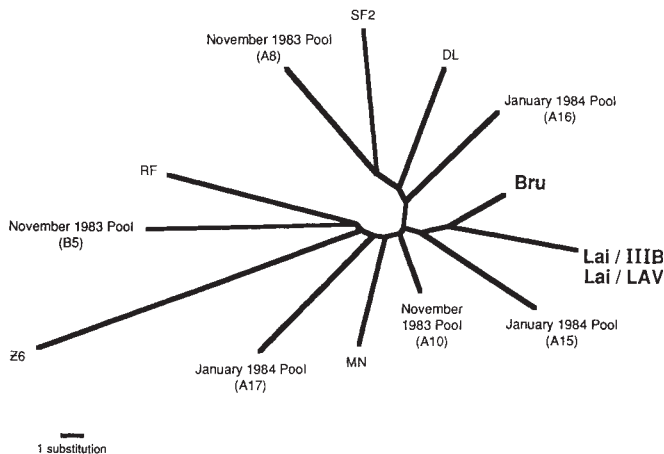


FIG. 1 Evolutionary tree contrasting the similarity of HIV-1 Lai/LAV and HIV-1 Lai/IIIB with their differences from variants of other HIV-1 isolates. This unrooted evolutionary tree is based on nucleotide substitution differences among 14 *env* V1/V2 sequences of variants of HIV-1 isolates: the major variant of each of the seven novel HIV-1 isolates reported in this study, as well as HXB2 of Lai/IIIB (pool and M2T-/B), J19 of Lai/LAV (patient LAI), JBB/LAV of Bru (patient BRU), and Los Alamos database sequences⁵ of variants of North American/Caribbean isolates (MN, SF2, RF), and an African isolate (Z6). Branch lengths are proportional to the inferred number of substitutions estimated to have taken place within the alignable sequence regions since any pair of sequences shared a common ancestor. The degree of difference between two sequences is illustrated by the combined length of the branches connecting them. Among this group of distantly related sequences, the most variable parts of the *env* V1/V2 regions cannot be aligned with confidence, and these regions must be omitted from evolutionary analysis. In the alignable regions, HXB2 of Lai/IIIB and J19 of Lai/LAV are identical; therefore on this tree they form a single, unforked branch. Sequences were aligned by hand using ESEE 2.00a. Alignments are available on request. In constructing the tree, 186 sequence positions (6,583–6,619, 6,683–6,775, 6,794–6,797 and 6,800–6,851 of HXB2⁵) were used. Of these, 75 contained a substitution in one or more of the sequences. The branching order and branch lengths were determined by the Neighbor-joining algorithm²⁷ using the computer program NJTREE, and are based on pairwise substitution differences among the sequences, without application of a multiple-hit correction. Bootstrap analysis does not strongly support the grouping of any of the major branches on this tree; therefore, the particular order in which the isolates evolved from each other cannot be inferred from these data. NJTREE produces unrooted trees; that is, trees in which the position of the earliest divergence, and therefore the direction of time, are not implied. Unrooted trees can be subsequently rooted based on biological knowledge of which branch represents the earliest divergence. Parsimony bootstrapping also failed to support any association of branches significantly. GenBank accession numbers for the novel sequences in this figure are M96456 to M96462. Four additional sequences from sample A10 and three from sample B5 that illustrate the variability within a sample were deposited in GenBank, with accession numbers M96493 to M96496, and M96497 to M96499, respectively.

from a single patient, and is also supported by the isolation of the variant NL43¹¹ from a culture derived from patient LAI. The degree of variability in the *env* V1/V2 region within a single patient observed in some of the samples (such as A10 and B5), and within each of two recent clinical samples, was greater than that between variants of HIV-1 Lai/LAV and HIV-1 Lai/IIIB (data not shown).

The rate of nucleotide substitution within the hypervariable regions of the *env* gene has been estimated to be within a range of 0.4% to 1.6% per site per year^{15–18}. Other studies have reported identical or closely related sequences among variants from different patients only if the samples were collected shortly after the infection and if the infections were related, that is, through transfusions with the same contaminated lot of blood or through mother–infant transmission^{16,19,20}. Therefore, for identical quasiespecies of HIV-1 Lai/LAV and HIV-1 Lai/IIIB to have

been derived independently from two different patients would require that the infections were related and that samples from both patients had been taken very shortly after their common infectious event. LAI was an AIDS patient with Kaposi's sarcoma when the initial sample was taken⁶, indicating a much earlier infection date. Thus patient LAI could not have been recently infected by one of the patients who contributed samples to LTCB. If the source of HIV-1 Lai/IIIB was not patient LAI, it could only be a person directly or indirectly infected by patient LAI very shortly before samples were obtained at the Pasteur Institute and LTCB. Therefore, it is highly implausible that HIV-1 Lai/LAV and HIV-1 Lai/IIIB were derived independently from two different individuals. It appears most likely that both the pool and the MoV culture were contaminated by variants of HIV-1 Lai when the M2T-/B culture was used to infect various cell types at LTCB⁷. □

Received 6 July 1992; accepted 22 March 1993.

1. Barré-Sinoussi, F. *et al.* *Science* **220**, 868–871 (1983).
2. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9–17 (1985).
3. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497–500 (1984).
4. Ratner, L. *et al.* *AIDS Res. Hum. Retrovir.* **3**, 57–69 (1987).
5. Myers, G., Rabson, A. B., Berzofsky, J. A., Smith, T. F. & Wong-Staal, F. *Human Retroviruses and AIDS* (Los Alamos National Laboratory, Los Alamos, NM, 1991).
6. Wain-Hobson, S. *et al.* *Science* **252**, 961–965 (1991).
7. Guo, H.-G. *et al.* *Nature* **349**, 745–746 (1991).
8. Goodenow, M. *et al.* *J. AIDS* **2**, 344–352 (1989).
9. Ratner, L. *et al.* *Nature* **313**, 277–283 (1985).
10. Muesing, M. A. *et al.* *Nature* **313**, 450–458 (1985).
11. Adachi, A. *et al.* *J. Virol.* **59**, 284–291 (1986).
12. Saiki, R. K. *et al.* *Science* **230**, 1350–1354 (1985).
13. Saiki, R. K., Walsh, P. S., Levenson, C. H. & Erlich, H. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6230–6234 (1989).
14. Schüpbach, J. & Popovic, M. *Science* **224**, 503–505 (1984).
15. Hahn, B. H. *et al.* *Science* **232**, 1548–1553 (1986).
16. Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. *J. Virol.* **64**, 6221–6233 (1990).
17. Wolfs, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 9938–9942 (1990).
18. Li, W.-H., Tanimura, M., & Sarp, P. M. *Molec. Biol. Evol.* **5**, 313–330 (1988).

19. Burger, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 11236–11240 (1991).
20. Wolinsky, S. M. *et al.* *Science* **255**, 1134–1137 (1992).
21. Kellogg, D. E. & Kwok, S. in *PCR Protocols: A Guide to Methods and Applications* 337–347 (Academic, San Diego, 1990).
22. Longo, M. C., Berninger, M. S. & Hartley, J. L. *Gene* **93**, 125–128 (1990).
23. Kwok, S. Y. in *PCR Protocols: A Guide to Methods and Applications* 142–145 (Academic, San Diego, 1990).
24. Kawasaki, E. S. in *PCR Protocols: A Guide to Methods and Applications* 146–152 (Academic, San Diego, 1990).
25. Chang, S.-Y., Shih, A., & Kwok, S. in *Molecular Evolution: Producing the Biochemical Data* (Academic, Orlando, in the press).
26. Erlich, H. A., Gelfand, D., & Sninsky, J. J. *Science* **252**, 1643–1651 (1991).
27. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).

ACKNOWLEDGEMENTS. We thank J. Hallum, S. Hadley and C. Watkins, OSI of NIH; W. Blattner, R. Gallo, T. Mays, B. Hahn, M. Martin, L. Montagnier, G. Schochetman, R. Ting and G. Curt for samples; J. Sninsky and S. Kwok for their advice in design of the PCR experiments and, with D. Gelfand and H. Erlich, for their review of the manuscript; R. Gibbs and S. Wolinsky for providing manuscripts on HIV-1 transmission between related patients before publication; T. Speed, M. Stoneking and W. K. Thomas for helpful discussion; and E. L. Cabot for an advance copy of ESEE2.00a.

Three-dimensional structure of photosystem II

Andreas Holzenburg^{*†}, Maria C. Bewley^{*},
Fiona H. Wilson^{*}, William V. Nicholson[‡]
& Robert C. Ford[‡]

^{*} Department of Biochemistry & Molecular Biology, and

[†] Department of Genetics, University of Leeds, Leeds LS2 9JT, UK

[‡] Department of Biochemistry and Applied Molecular Biology, UMIST,
PO Box 88, Manchester M60 1QD, UK

THE photosynthetic oxidation of water occurs within a multi-protein complex called photosystem II (PSII), located in the thylakoid membranes of plants and cyanobacteria. Little information on PSII architecture is available; even the oligomeric nature of the complex is contentious, with various predictions of monomeric¹, dimeric and tetrameric^{2–4} forms. Biochemistry suggests that it consists of an outer shell of easily removable light-harvesting proteins and a more resistant core⁵. The core itself can be split into further light-harvesting proteins and a reaction centre^{5,6}. The reaction centre polypeptides bind the electron transfer components, and probably contain amino-acid residues liganded to the cluster of four Mn atoms that is believed to constitute the water 'splitting' site^{6–8}. We report here the first, to our knowledge, three-dimensional functionally intact structure of PSII of

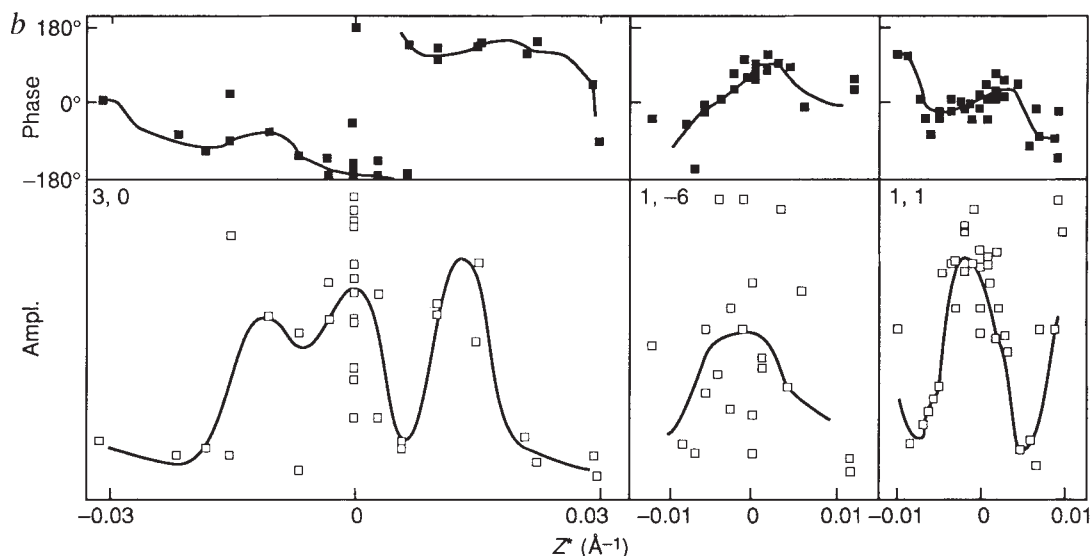
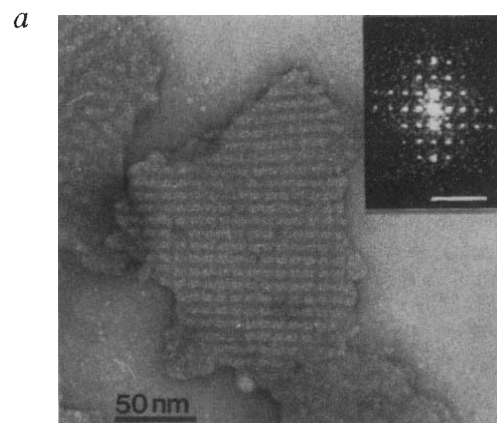
higher plants, complete with antennae and core light-harvesting proteins. Our model has been obtained by digital image processing of ordered two-dimensional arrays of the complex, such as have often been observed^{4,9–12}.

Figure 1a shows the appearance of ordered two-dimensional arrays of PSII in spinach chloroplast membranes. Surrounding the regular array are other membranes containing randomly oriented particles. In other cases we observed a mixture of nonordered and ordered particles within the same membrane with no apparent difference in size and shape, as previously reported^{10,12}. The identity of the complexes within the ordered arrays was confirmed by analysis of the polypeptide composition of the various membrane fractions by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting with antibodies raised against the PSII polypeptide D2 (Fig. 2).

Digital image analysis of the ordered two-dimensional arrays was used to obtain an averaged projection of the unit cell¹³. By combining the data from a series of micrographs recorded at various angles relative to the electron beam, a three-dimensional map was calculated. Figure 1b shows amplitudes and phases along different layer lines. Information about the missing data lost by the impracticality of tilting the specimen to 90° was obtained from side-on projections using: folded arrays; by plotting density against section number in the three-dimensional map; and also by using low-angle X-ray scattering data¹¹ and electron microscopy data of thin-sections¹⁰. To minimize interpretation problems, all the features discussed originate well

FIG. 1 a, An electron micrograph of negatively stained¹³ PSII-enriched membranes¹⁹. The inset shows the power spectrum of a selected area of the two-dimensional array with strong reflections up to the 4th order. Unit cell dimensions were $a=16.8$ nm, $b=18.9$ nm with $\gamma=91^\circ$ and $p1$ plane group symmetry. Weak reflections up to the 6th order were recorded, corresponding to a spatial resolution of 3 nm; scale bar, 0.2 nm^{-1} . b, Variations of amplitudes and phases with z^* along lattice lines (3, 0) (1, -6) and (1, 1). Maximum relative amplitudes shown correspond to 210, 25 and 107, respectively.

METHODS. Three-dimensional data were obtained from 39 films of two-dimensional arrays, each taken at a calibrated magnification of $\times 27,700$, comprising a total of 25 independent equally spaced projections at nominal tilt angles between $+60^\circ$ and -60° . For each film, amplitudes and phases including 6th order reflections were refined. Tilt angles and axes were calculated using the ratio of a^* and b^* and the included angle according to ref. 20. The phases of all films within the first data set were then refined into a common origin using the program ORIGIN (MRC, Cambridge). Data sets were then either (1) merged to form the first (core) data set, or (2) merged independently and then combined with the core data by subsequent sorting. Using 1,860 measured reflections, with an average phase residual of 27.8° , smooth curves were fitted along each lattice line sampled at $1/20$ nm. The corrected amplitudes and phases were used to calculate a three-dimensional Fourier map from 414 unique reflections. The specimen thickness was evaluated as described in the text, suggesting a maximum overall thickness of $z=10$ nm. Initially a specimen thickness of 20 nm was used during data merging, and this was then narrowed to 15 nm for calculation of the three-dimensional Fourier map and adjusted to 10 nm for the final presentation. Data were analysed on a VAX 4000/60 workstation using the MRC Cambridge image processing system and displayed within the framework of the FRODO 4.4 software package.



within the limits of the three-dimensional map. In Fig. 3a-c, part of the map of the unit cell is presented by selected slices taken parallel to the membrane plane. The three-dimensional structure presented in Fig. 4 depicts stereo pairs of the unit cell. The volume enclosed by the lowest contours ($1,100 \text{ nm}^3$) would correspond to a protein complex of M_r 890K (using a partial specific volume of $0.74 \text{ cm}^3 \text{ g}^{-1}$), which agrees with the prediction of a single PSII complex per unit cell.

In electron micrographs of negatively stained specimens, the data represents the positions of heavy atoms constituting the stain, with protein represented by stain-excluding areas. Despite a generally observed resolution limit of about 2 nm, negative stain is considered to be useful for distinguishing the arrangement of subunits within a large protein complex because stain can partially penetrate into it, but tends to be excluded from the hydrophobic cores of subunits¹⁴. In negative stain, phase errors are smaller than those for the amplitudes (see Fig. 1b), and this leads to more accuracy on the position rather than the volume of domains.

Previous studies of PSII-containing membranes have identified the inward and outward facing sides of the complex^{10,12}. We can thus identify the two sides of the complex with confidence (Figs 3 and 4). The location of the 4.5-nm-thick membrane plane is weakly defined in the three-dimensional structure, partly because of the lower resolution perpendicular to the membrane plane, and perhaps because of some stain penetration into the membrane. By plotting contrast levels against horizontal section number, a central minimum at sections 3 to 4 (Fig. 3) was observed which may correspond to the central section of the lipid bilayer. Note the asymmetry of the complex across the membrane which was suggested by others^{10-12,15-17}.

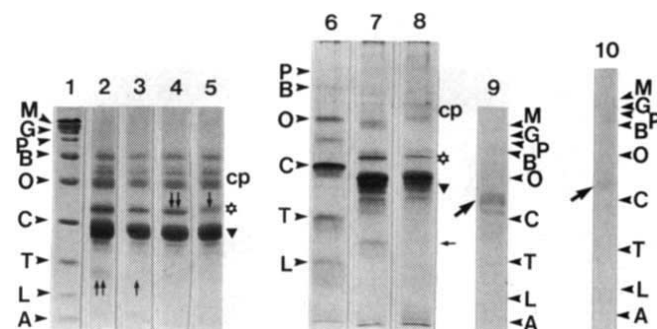


FIG. 2 Gel electrophoresis of PSII proteins²¹ using 12% (w/v) acrylamide resolving gels. Samples were mixed 1:1 with 0.13M Tris-HCl, 6% (w/v) SDS, 40% (v/v) glycerol, 10 mM dithiothreitol, 20 mg ml^{-1} bromophenol blue, pH 6.7, and incubated at room temperature for 2 h. Tracks 1 and 6 contained M_r markers of 200K (M), 116K (G), 97K (P), 66K (B), 45K (O), 31K (C), 21K (T), 14K (L) and 6.5K (A). Tracks 2-5 contained PSII-enriched membranes¹⁹ equivalent to 7 μg chlorophyll and were Coomassie stained. Track 2, Membranes similar to those used for structural analysis. Track 3, Prewashed with 0.3 M sucrose, 0.3 M NaCl. Track 4, Prewashed with 0.3 M sucrose, 1 M NaCl. Track 5, Prewashed with 0.3 M sucrose, 0.8 M Tris (pH 8.2). These tracks show the typical polypeptide pattern of PSII-enriched grana membranes with the LHCP the strongest band at 27-29K (triangle) and the core light harvesting proteins CP47 and CP43 (cp). Extrinsic proteins of 33K (star) and 17K (arrows) are removed as expected by the various salt- and Tris-washing procedures. The 23K extrinsic polypeptide is not well resolved in this gel. Tracks 7 and 8 contained material equivalent to 0.14 μg chlorophyll and were silver stained²². Track 7, Untreated membranes. Track 8, Membranes pre-washed with Triton X-100 at a detergent:chlorophyll ratio of 40, which yields material containing 60% two-dimensional arrays as judged by electron microscopy. Western blotting (tracks 9, 10) was done according to ref. 23 with the primary antibody directed against the D2 subunit of PSII at dilutions of 1:100 (track 9) and 1:400 (track 10). Track 9, Untreated membranes. Track 10, Material containing 60% two-dimensional arrays. A band of 32-34K reacted with the antiserum in both tracks (Thick arrows).

The internal or luminal face is characterized by four strong central domains (luminal domains I-IV, Fig. 3) surrounding a cavity. The identity of these four domains remains in doubt, although it is likely that the hydrophilic polypeptides of the oxygen-enhancing complex¹⁸ could form at least one of them. The pseudo-twofold or -fourfold symmetry of the domains has previously given rise to speculation that a dimeric or tetrameric PSII complex may exist²⁻⁴; this can now be discounted because no such symmetry is found in the three-dimensional structure (Fig. 4), and because of the mass of the complex, as discussed above. On the periphery, eight evenly spaced protrusions are discernible (Fig. 3, numbered 1-8) which are less contrasted than the four major domains and are incompletely resolved on the luminal side. Although contacts along the rows of PSII molecules are apparent, there is no indication of any contact between the rows. The model shows that there is little variation of the overall structure on the luminal face. The high contrast is due to stain trapping between the complexes and the specimen support film.

On the outer or stromal face of the complex, the stain is shallower. Eight protrusions are resolved and six are relatively easy to assign to their corresponding positions on the luminal side of PSII (Fig. 3). Four of them (1, 2, 5, 6) appear to be involved in making contacts between the rows of molecules in the arrays. For the following reasons, we suggest that these

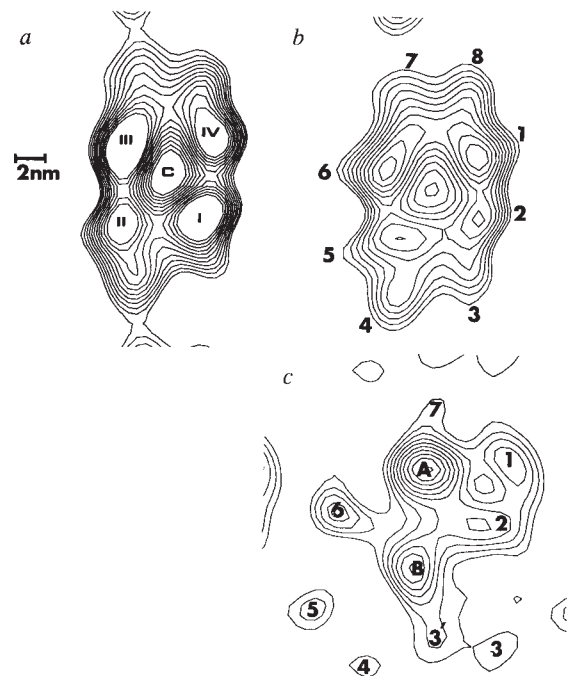
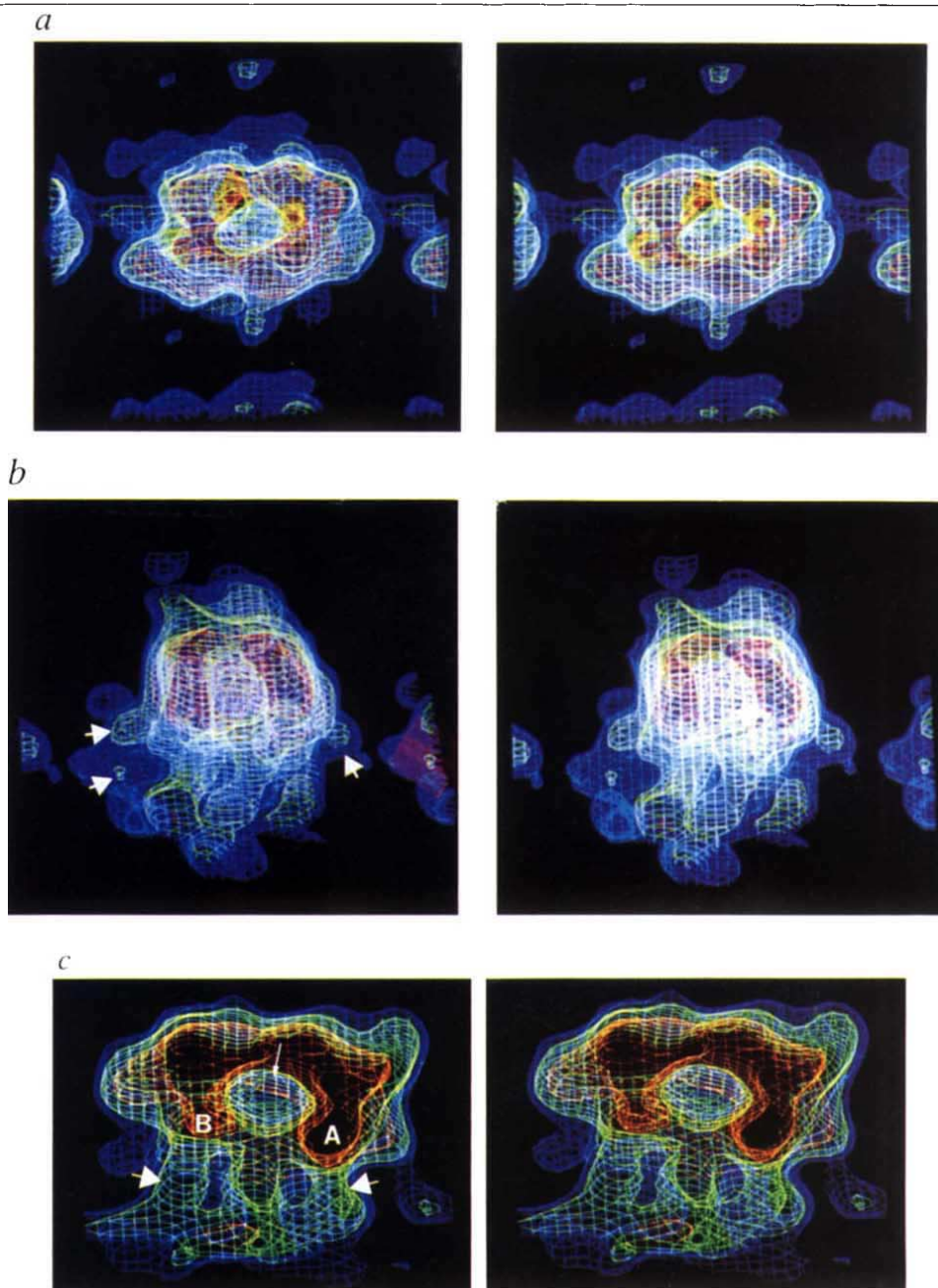


FIG. 3 Selected sections from the three-dimensional model which was plotted as 21 sections (-10 to 10) taken parallel to the membrane plane and spaced 0.5 nm apart, with section 0 at the centre of the unit cell. Sections -5 (a), 0 (b) and 5 (c) are shown with the viewing direction from the outer (stromal) side of the membrane. Contour lines begin at the mean density level (zero) and extend up to 250 in units of 25. The maximum contour line is equivalent to a density 3.5 times the standard deviation above the mean level. The luminal side of the complex shown in a- and b is characterized by 4 large domains (I-IV) and 8 smaller protrusions (1-8) surrounding a central cavity (C). The stromal side shown in c consists of a central ridge composed of two domains A and B. The ridge is surrounded by a ring of smaller features (1-7) that appear to emerge from the luminal protrusions (1-7). Luminal protrusion 8 cannot easily be assigned to a stromal feature at present. Both stromal features 3 and 3' emerge from the luminal protrusion 3. Connections between unit cells across the rows of PSII complexes (that is along a-axis) are formed by features 1, 2 and 5, 6. Along the rows, connections are made by features 4 and 7. The a- and b-axes are horizontal and vertical, respectively.

FIG. 4 Stereo pairs representing different views of the PSII complex contoured at 0.5, 1 and 2 times the standard deviation above the mean density level (blue, yellow and red netting, respectively). *a*, The membrane is perpendicular to the viewing direction which is from the luminal side. The *a*- and *b*-axes are vertical and horizontal, respectively. Protrusion 6 is at the bottom of the picture, with 1 and 2 at the top and 7 on the right. *b*, The complex is shown end-on (parallel to *b*-axis), viewed from the luminal side and slightly above the membrane plane. The contacts between the rows on the stromal side can be seen with protrusions 1 and 2 on the left and 6 on the right of the picture (arrows). *c*, The complex as in *a*, but tilted around the *b*-axis to obtain a luminal view from slightly above the membrane plane. The nearest part of the complex has been sliced away to reveal the interior with the four luminal domains merging into the two stromal domains A and B. The round shape in the centre is the stained-filled cavity (thin arrow) which extends up to the luminal surface. Note that the protein densities of domain A and B extend towards the stromal face. The large arrows point to the expected position of the outer (stromal) leaflet of the lipid bilayer.



features may represent monomeric light-harvesting chlorophyll *a/b* proteins (LHCPs): (1) their location on the periphery; (2) their number (8–10 LHCPs would bind about 120–150 chlorophyll molecules); and (3) their bias towards the stromal face¹⁰. Within this scaffolding of putative LHCPs lie two prominent domains about 5 nm in diameter. Domain A emerges from

luminal domains I, III and IV, and domain B is continuous with luminal domain II. It is likely that one of these domains will be the reaction centre of PSII, which bears homology with the bacterial photosynthetic reaction centre⁸. The presented model should help to establish investigations into structure–function relationships of PSII. □

Received 23 December 1992; accepted 22 March 1993.

1. Dekker, J. P., Betts, S. D., Yocum, C. F. & Boekema, E. J. *Biochemistry* **29**, 3220–3225 (1990).
2. Haag, E., Irrgang, K. D., Boekema, E. J. & Renger, G. *Eur. J. Biochem.* **189**, 47–53 (1990).
3. Peter, G. F. & Thornber, J. P. *Pl. Cell Physiol.* **32**, 1237–1250 (1991).
4. Bassi, R., Magaloli, A. G., Tognon, G., Giacomelli, G. M. & Miller, K. R. *Eur. J. Cell. Biol.* **50**, 84–93 (1989).
5. Dainese, P. & Bassi, R. *J. Biol. Chem.* **266**, 8136–8142 (1991).
6. Nanba, O. & Satoh, K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 109–112 (1987).
7. Nixon, P. J. & Diner, B. A. *Biochemistry* **31**, 942–949 (1992).
8. Eisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318**, 618–624 (1985).
9. Park, R. B. & Biggins, J. *Science* **144**, 1009–1011 (1964).
10. Simpson, D. J. *Carlsberg Res. Commun.* **44**, 305–336 (1979).
11. Kreutz, W. *Adv. Bot. Res.* **3**, 54–99 (1970).
12. Staehelin, L. A. *Biochim. biophys. Acta* **408**, 1–11 (1975).
13. Holzenburg, A., Wilson, F. H., Finbow, M. E. & Ford, R. C. *Biochem. Soc. Trans.* **20**, 591–597 (1992).
14. Amos, L. A., Henderson, R. & Unwin, P. N. T. *Prog. Biophys. molec. Biol.* **39**, 183–231 (1982).

15. Brandt, J., Nielsen, V. S., Christensen, H. T., Simpson, D. J. & Økkel, J. S. *Pl. molec. Biol.* **19**, 699–703 (1992).
16. Shinozaki, K. *et al. EMBO J.* **5**, 2043–2049 (1987).
17. Kühlbrandt, W. & Wang, D. N. *Nature* **350**, 130–134 (1991).
18. Ghanotakis, D. F., Babcock, G. T. & Yocum, C. F. *FEBS Lett.* **167**, 127–130 (1984).
19. Ford, R. C. & Evans, M. C. W. *FEBS Lett.* **160**, 159–163 (1983).
20. Shaw, P. J. & Hill, G. J. *Micron* **12**, 279–282 (1981).
21. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
22. Morrissey, J. H. *Analyt. Biochem.* **117**, 307–310 (1981).
23. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).

ACKNOWLEDGEMENTS. We thank R. Henderson of the MRC Laboratory of Molecular Biology, Cambridge for donating the MRC image analysis software and for training us in its use. A. Trebst for antibodies, C. Ellis, P. McPhie, S. Möller and L. Child for technical assistance, and D. N. Perkins for computational assistance and software development. We acknowledge grants from the Royal Society (A.H., R.F.), the Nuffield Foundation (A.H.) and the UK Agricultural and Food Research Council (A.H.).

Taking the bull by the horns

Genetic analysis of domestic animals demonstrates the feasibility of marker-assisted selection of economically important traits and may foreshadow good things to come (other than better bacon) for human geneticists.

"MOLECULAR genetics has arrived on the farm", concluded James Womack in a *Nature News and Views* article¹ late last year. So it has, there being several reasons behind the endeavour to understand the molecular genetics of domestic animals. Veterinarians wish to define the hereditary component of common diseases and thus eliminate them (negative selection); breeders aim to identify and exploit markers of economically important characteristics of their animals (positive selection); and geneticists can investigate genome organization and evolution by examining homologous sequences across species (comparative mapping).

Whereas comparative mapping requires extensive and precise genome maps from several species (see ref. 2 for review), isolating markers of diseases or desirable traits can, and is, progressing piecemeal. In the June issue of *Nature Genetics*, groups led by Michel Georges and Jay Hetzel³ describe a linkage study in which the autosomal *polled* locus (controlling horn development) is seen to segregate with two microsatellite markers on chromosome 1 in *Bos taurus* (a classification that includes most of the well-known European cows). Previously it was only assumed that a *polled* locus was the main determinant of the horn phenotype⁴. Georges *et al.* confirm both the existence of the locus and its mode of inheritance, paving the way for generation of a physical map of the region and isolation of the gene. It can be assumed that, before the gene itself is isolated, closer markers (there is an estimated 13 per cent recombination rate between the two microsatellite markers and the *polled* locus) will be found allowing marker-based selection of the trait.

Why make a fuss about a rudimentary linkage study? Horns are generally considered undesirable (they present a risk of injury to both the animals and their handlers) and currently a surgical procedure is routinely used to de-horn cattle. But although the possibility of being able to

breed hornless animals is enticing, the real interest in this work is that it demonstrates the feasibility of marker-based selection in domestic animals.

Until now, breeders searching for an improved characteristic in their animals have had to embark on time-consuming breeding programmes. Despite the limitations of this approach, the economic potential of a slightly improved animal is such that extensive and elaborate schemes involving thousands of animals have been established. Molecular geneticists offer an alternative, and although breeders have shown some scepticism towards the new genetics impatience with traditional approaches (which can take 20–30 years to establish a new breed) has forced them to sit up and pay attention.

Four years ago the first bovine reference pedigrees (directly comparable to the indispensable human pedigrees produced by the Centre d'Etude du Polymorphisme Humain) were established by the Commonwealth Scientific and Industrial Research Organization in Queensland, Australia. This group and others (notably Genmark of Salt Lake City) started generating genetic markers. Their greatest challenge, the same as that of human geneticists, is to establish evenly spaced genome markers and use them to define single-gene traits and (ultimately) unravel the complicated genetics of multifactorial traits. Just as human geneticists have struggled with very sparse linkage maps, so too have animal geneticists, and with the publication of the first linkage studies that convincingly establish (distant) markers to single-gene traits^{3,5} they are reaping the rewards.

This work is now well advanced. The cytogeneticists who are able to identify the 30, mostly acrocentric, chromosomes of the bovine genome are using *in situ* hybridization to map sequences. An international collaboration is working on a bovine genome map, and linkage maps for chromosomes 1 and 21, with markers spaced at 20 centimorgans, are now complete. This same collaboration has also established large full-sibling pedigrees to complement the existing reference pedigrees, and is pooling somatic cell hybrids. Animal breeders hope to be able to select characteristics such as improved milk production, better disease resistance, greater muscle mass and reduced fat content, and genetically to select for flavour and tenderness of meat. On the face of it these

are daunting aims; nonetheless, certain breeds do produce more tender meat and some breeders would even claim to be able to differentiate between breeds on the basis of meat flavour.

The repercussions of this enterprise go further, for as domestic animal genes are identified and sequenced they will contribute to the data on comparative mapping. Indeed comparisons between sheep and cattle have shown that about half of the cattle markers may be useful in sheep studies (a not inconsiderable discovery, as very few markers are established in sheep linkage maps⁶). Furthermore, there is the prospect of producing a new generation of domestic animal models of human genetic diseases. Those massive breeding programmes, underway for many years, have thrown up a variety of unusual phenotypes. Although most such animals are slaughtered early in life, some are of great interest to the farming community (for instance, in contrast to the human condition, muscle hypertrophy in an animal reared for its meat is a very valuable characteristic). Occasionally, a particular phenotype will be recognized as bearing similarities to a human disease and the animal will be further investigated. Bovine models for Marfan's syndrome (for which no other animal model has been reported), Pompe's disease (glycogen storage disease II), osteogenesis imperfecta and Ehlers-Danlos syndrome have already been identified, and although there are drawbacks to having such a large animal as a model there are also some benefits — like humans, for instance, the cow has a nine-month gestation period, and the size of the animals means that there is no shortage of material for investigating fetal development.

It will be some time before complete comparative maps, well-characterized models or selection of multifactorial traits become commonplace. But the paper by Georges *et al.* is one of those straws which show which way the wind is blowing and makes plain that these are worthwhile and realistic goals.

Adrian J. Iverson

Adrian J. Iverson is assistant editor of Nature Genetics.

Also in this month's *Nature Genetics*: β -cGMP phosphodiesterase mutations in recessive retinitis pigmentosa; an artificial neo-organ for gene therapy to correct Sly syndrome; novel gene fusion event produces tumour-specific marker; and screening for unstable trinucleotide repeats.

1. Womack, J. E. *Nature* **360**, 108–109 (1992).
2. O'Brien, S. *et al.* *Nature Genet.* **3**, 103–112 (1993).
3. Georges, M. *et al.* *Nature Genet.* **3**, 206–210 (1993).
4. Long, C. R. & Gregory, K. E. *J. Hered.* **69**, 395 (1978).
5. Georges, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **90**, 1058–1062 (1993).
6. Montgomery, G. W. *et al.* *Nature Genet.* (in the press).

Sequence-specific scission of DNA by RNAs linked to a chemical nuclease

David S. Sigman, Chi-hong B. Chen and Michael B. Gorin

RNAs linked to the chemical nuclease 1,10-phenanthroline-copper cut double-stranded DNA of complementary sequence. This cleavage reaction is applicable to all sequences and can be used to measure the distance between marker genes in base pairs, map the size of a transcription unit and define positions of chromosomal breakpoints.

THE chemical nuclease 1,10-phenanthroline-copper can be linked to enzymatically synthesized RNA in which 5-allylamine-UTP is used as the sole source of UTP. The derivatized RNA retains all the hybridization properties of RNA and forms R-loops with complementary DNA sequences. Following the formation of these three-stranded structures, which rely on Watson-Crick base-pairing for recognition, scission can be activated by the addition of copper ion, thiol and hydrogen peroxide. Applications of this new method of DNA scission would include measurement of the distance between two DNA sequences in base pairs, mapping the size of a transcription unit using sequences derived from a cDNA and identifying the nucleotide sequences at chromosomal breakpoints¹. The strength and novelty of the method is that no special sequence is required for the formation of a triple helix² or recognition by a restriction enzyme or a DNA binding protein^{3,4}.

Chemical nucleases

'Chemical nucleases' are redox-active coordination complexes that cleave nucleic acids by oxidative attack on the (deoxy)ribose moiety of DNA and RNA⁵. The first reagent of this type was the 2:1 1,10-phenanthroline-cuprous complex [(OP)₂Cu⁺], which cleaves DNA using hydrogen peroxide as a co-reactant by oxidatively attacking the deoxyribose moiety at the C-1 hydrogen leading to a series of elimination reactions that ruptures the phosphodiester backbone and yields 3' and 5' phosphomonoester termini, free bases and 5-methylene-furanone⁶⁻⁸ (see Fig. 1).

The free coordination complex, like pancreatic DNase I, does not cleave DNA equivalently at all sequence positions. The major determinant in its scission preference is the base pair associated with the nucleotide 5' to the site of scission. However, the chemical nuclease activity of (OP)₂Cu⁺, and other

redox-active coordination complexes, can be targeted by linking them to deoxy-oligonucleotides, proteins and organic ligands that bind to nucleic acids (for example, distamycin and Hoechst 33258)⁹⁻¹¹. For example, 1,10-phenanthroline linked to deoxyoligonucleotides cleaves single-stranded DNA and RNA with high sequence selectivity following heat denaturation and reannealing^{12,13}. Three proteins, the *Escherichia coli* trp repressor¹⁴ and catabolite activating protein (CAP)^{15,16} and the lambda phage cro proteins¹⁷ have been converted into site-specific scission reagents by covalent modification with a 1,10-phenanthroline derivative. The chimaeras prepared with the trp repressor and the CAP protein are especially noteworthy because they are very efficient in making double-stranded breaks in their target sequences. Effectively, they are 'restriction' enzymes with 15-bp recognition sequences. When linked to a carrier ligand only one phenanthroline/copper is essential for cleavage. Unlinked to a carrier ligand a second phenanthroline is necessary to increase the affinity of the coordination complex for DNA.

R-loop directed scission

RNAs can also be used as carriers for the nucleolytic activity of 1,10-phenanthroline-copper and, in principle, direct scission at any nucleic acid sequence. Moreover, since 5-allylamine-UTP¹⁸ is accepted as a substrate by T-7 and SP-6 RNA polymerase, RNAs with a pendant reactive allylamine at all uridines can be readily synthesized using readily available vectors or synthetic promoters. A simple protocol using the chemical cross-linking reagent, *N*-succinimidyl 3-(2-pyridyldithio)propionate, can then be used to link 1,10-phenanthroline to the RNA at internal positions within the sequence (1) (see Fig. 2). The availability of multiple sites of modification leads to a heterogeneously derivatized RNA but also increases the probability that any hybridized RNA will cleave the target nucleic acid. These chemically modified RNAs cut complementary single-stranded DNAs very efficiently¹.

However, to cleave double-stranded DNA, the duplex must be partially melted so that the RNA can recognize its complementary sequence. This can be

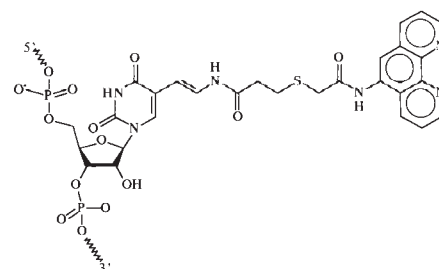


FIG. 2 RNA derivatized internally with 1,10-phenanthroline.

achieved through the formation of R-loops, which are three-stranded structures in which all recognition involves Watson-Crick base pairing. They were initially used to map transcription units using electron microscopy¹⁹ and to select low abundance mRNA species²⁰. R-loops typically form in 70 per cent formamide at 55 °C although conditions vary with the GC content of the region. Although formamide is essential for their formation, R-loops are kinetically stable and can be isolated by chromatography or by electrophoresis. Recent studies indicate that R-loops prepared from 40-nucleotide long RNAs are stable to electrophoresis. Addition of the essential co-reactants cupric ion, 3-mercaptopropionic acid, and hydrogen peroxide to the solution or gel slice containing the R-loop results in scission, which can be efficiently quenched with 2,9-dimethyl-1,10-phenanthroline or EDTA. The products of the cleavage reaction can be separated in a second electrophoretic dimension; yields for double-stranded scission hover in the range of 5–10 per cent for a plasmid test system.

Although the chemistry of double-stranded cleavage is not robust, low yields may be tolerable for the determination of the distance between two marker genes in base pairs. In this type of experiment (Fig. 3), RNAs derived from the sequence of two marker genes are prepared and then derivatized with 1,10-phenanthroline. Two R-loops are formed simultaneously with the OP-modified RNAs (RNA-OP). After scission, the products are then subjected to pulsed-field gel electrophoresis and visualized by Southern blotting using segments of the marker genes as probes. The size of the intervening fragment can be measured in an appropriately calibrated gel (see Fig. 3). In initial experiments, the delta crystallin I and

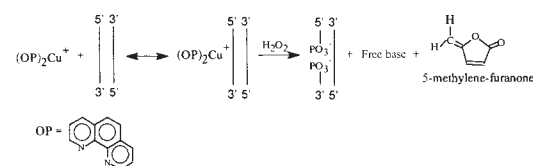


FIG. 1 Mechanism of DNA cleavage by 1,10-phenanthroline-copper.

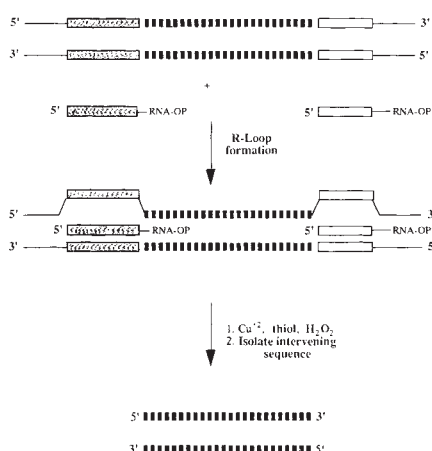


FIG. 3 Method for determining the distances between two probe sequences. RNA-OP:OP-modified RNAs.

2 locus of chick DNA was used as a test system. Two nonallelic crystallin genes are separated by roughly 4.2 kbp and arranged in the same transcriptional polarity^{21,22}. One RNA was derived from the 5' end of the delta1 sequence and the second from the 3' end of delta2. RNA from these regions was prepared using Bluescript vectors and modified with 1,10-phenanthroline. A new 24-kbp DNA fragment was evident following digestion of chick genome DNA (estimated minimum size greater than 100 kbp). This corresponds to the size anticipated from the known sequence of this locus. Although the yield of the fragment based on hybridizable counts was only 10 per cent, more efficient scission is not necessary if the goal is simply to hop along a chromosome in roughly 25-kbp bounds. No attempt has yet been made to clone the fragments liberated.

A second potential application of this methodology is to map the genomic size of transcription units using sequences derived from cDNAs. Suppose RNAs corresponding to the sequence of the cDNA from positions 1–40 and 250–290 are synthesized and used to cleave genomic DNA. If only short fragments (about 200 bp) are isolated, no intron separates the sequences used in the targeted cleavage. In contrast, any fragment substantially larger than 200 bp indicates the presence and size of one or more introns.

A final application of this methodology that will be investigated will be the identification of chromosomal breakpoints. Selecting sequences from the two chromosomes involved in a balanced translocation would help identify more precisely the sequences that border the breakpoint.

Potential difficulties

The flexibility provided by RNA-directed scission via R-loops is its greatest advantage. However, two significant limitations of this method currently exist. The first is that the yield of double-stranded breaks always hovers in the 5–10 per cent range. Secondly, standard agarose used for *in situ* lysis of whole cells to prepare intact chromo-

somal DNA cannot withstand the formamide concentrations required for R-looping. Attempts to resolve the first problem include tethering other chelating agents to the uridine residues or possibly developing strategies of linking staphylococcal nuclease to the hybridizing RNA²³. The second problem may be less of a concern than originally anticipated, especially if one is content with short hops of 25–75 kbp in length or uses solution methods capable of preparing DNA with an average fragment size of 500 kbp²⁴. In addition, the synthesis of agarose derivatives capable of withstanding R-loop conditions can also be envisioned.

Despite these apparent limitations, the programmable R-loop strategy is appealing and makes further investment to improve this methodology a worthwhile objective. Experiments are underway to test its utility in the applications outlined above. □

David S. Sigman and Chi-hong B. Chen are in the Department of Biological Chemistry, School of Medicine, Department of Chemistry and Biochemistry, and the Molecular Biology Institute, University of California, Los Angeles, California 90024–1570, USA. Michael B. Gorin is in the Departments of Ophthalmology and Human Genetics, University of Pittsburgh, 203 Lothrop Street, Pittsburgh, Pennsylvania 15213, USA. For more information, fill in reader service number 100.

- Chen, C.-h. B., Gorin, M. B. & Sigman, D. S. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4206–4210 (1993).
- Dervan, P. B. *Nature* **359**, 87–88 (1992).
- Koob, M. & Szybalski, W. *Science* **250**, 271–273 (1990).
- Ferrin, L. J. & Camerini-Otero, R. D. *Science* **254**, 1494–1497 (1991).
- Sigman, D. S. & Chen, C.-h. B. *A. Rev. Biochem.* **59**, 207–236 (1990).
- Sigman, D. S., Graham, D. R., D'Aurora, V. & Stern, A. M. *J. biol. Chem.* **254**, 12269–12272 (1979).
- Goyne, T. E. & Sigman, D. S. *J. Am. chem. Soc.* **109**, 2846–2848 (1987).
- Sigman, D. S. *Acc. Chem. Res.* **19**, 180–186 (1986).
- Dervan, P. B. *Science* **232**, 464–471 (1986).
- Chen, C.-h. B., Mazumder, A., Constant, J. F. & Sigman, D. S. *Bioconjugate Chem.* **4**, 69–77 (1993).
- Sigman, D. S., Bruce, T. W., Mazumder, A. & Sutton, C. L. *Acc. chem. Res.* **26**, 98–104 (1993).
- Chen, C.-h. B. & Sigman, D. S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7147–7151 (1986).
- Chen, C.-h. B. & Sigman, D. S. *J. Am. chem. Soc.* **110**, 6570–6572 (1988).
- Sutton, C., Mazumder, A., Chen, C.-h. B. & Sigman, D. S. *Biochemistry* **32**, 4225–4230 (1993).
- Ebright, R. H., Ebright, Y. W., Pendergast, P. S. & Gunasekera, A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2882–2886 (1990).
- Pendergast, P. S., Ebright, Y. W. & Ebright, R. H. *J. cell. Biochem.* **17**, 156 Abs. K031 (1993).
- Bruce, T. W., Wise, J., Rosser, D. S. E. & Sigman, D. S. *J. Am. chem. Soc.* **113**, 5446–5447 (1991).
- Langer, P. R., Waldrop, A. A. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6633–6637 (1981).
- Thomas, M., White, R. L. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2294–2298 (1976).
- Rosbash, M. *et al. Meth. Enzym.* **68**, 454–469 (1979).
- Nickerson, J. M. *et al. J. biol. Chem.* **260**, 9100–9105 (1985).
- Nickerson, J. M. *et al. J. biol. Chem.* **261**, 552–557 (1986).
- Corey, D. R., Pei, D. & Schultz, P. G. *Biochemistry* **28**, 8277–8286 (1989).
- Longmire, J. L., Albright, K. L., Lewis, A. K., Meincke, L. J. & Hildebrand, C. E. *Nucleic Acids Res.* **15**, 859 (1987).

ADVERTISEMENTS

Plant and Animal Tissue
HOMOGENIZER
and EMULSIFIER

- Disrupts and disperses small samples (1/2–100 ml)
- Both mechanical and ultrasonic action
- 5,000–30,000 rpm
- Built-in speed control
- Easy to clean
- Best price (\$485 complete)

The TISSUE-TEAROR™
Manufactured By:
BIOSPEC PRODUCTS • PH/FAX (918) 336-3363
P.O. BOX 722 • BARTLESVILLE, OKLAHOMA 74005

Reader Service No. 15

Synthetic Oligonucleotides

Rapid Quality Worldwide Service

£2.50 per base, no setup!

option of H.P.L.C. purification
biotinylation, fluorescent labelling
RNA Synthesis

phosphothioates, many other modifications

OSWEL DNA SERVICE

Department of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road,
EDINBURGH EH9 3JJ

Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No.18

Professional Custom Services

DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

V.G. Services 1200 Wonderland Rd S Bldg #7 Unit #2
London, Ontario Canada N6L 1A8
CALL/FAX: (519) 652-3758 ALT. FAX: (519) 652 0053

Reader Service No.32

PEPTIDE SYNTHESIS SUPPORTS

New polydimethylacrylamide &
polystyrene supports in bulk
quantities * High loadings
* Reproducible high quality

For further details, tel:

USA: 413/253 9554

UK: 0694 723581

Netherlands: **Polymer Laboratories**
03404-22508
SEPARATION SCIENCE DIVISION

Reader Service No.8

Peptide and oligonucleotide profile

Featured this week — an epitope mapping system, an instrument for the purification of proteins, peptides and glycoproteins from complex mixtures by continuous elution electrophoresis and a radiolabelled ligand for brain research.

VERSION 4.0 of the MacVector **sequence analysis software** program from International Biotechnologies includes primer prediction and selection capabilities for the polymerase chain reaction and sequencing applications (*Reader Service No. 101*). It is also available with AssemblyLIGN, a sequence assembly and editing program. Researchers wishing to amplify a DNA fragment may use MacVector's new analysis to search for appropriate primer pairs according to product size or the region in which they occur. The program calculates melting temperature and optimum annealing temperature using the 'nearest neighbour' method. It also lists the numbers of primers and primer pairs it has considered, accepted and rejected, and it lists the reasons for rejection so that researchers can adjust their parameters to expand or contract the list of possible primers accordingly. Acceptable pairs are listed with their sequence, length, T_m and G + C content along with the location, length, T_m and G + C content of amplified product. An optimum annealing temperature for the reaction is also calculated. To round out its enhanced support for DNA sequencing, version 4.0 also allows users to screen sequences for DNA sequencing primers or hybridization probes. Oligonucleotide T_m and T_d calculations have been added to the base composition analysis.

AMS Biotechnology announces the availability of a rapid **epitope mapping system** from Novagen (*Reader Service No. 102*). The kit allows the researcher to determine epitopes for monoclonal antibodies, scan for epitopes using cDNA clones, map functional domains by affinity screening, scan viral/bacterial genomes for antigenic protein determinants, localize receptor binding sites, and produce peptides and pre-conjugated immunogens.

New on the market from Stratech Scientific is a reagent kit designed to provide an effective and rapid **screening method for the crystallization of macromolecules** (*Reader Service No. 103*). The company says that the screen provides a simple and practical means for finding initial crystallization conditions for proteins, peptides and oligonucleotides. By using sparse matrix sampling, the kit allows the user to test rapidly wide ranges of pH, salts and precipitants using a very small sample of macromolecule. The kit contains 50 reagents of varying salt and precipitant composition over five pH levels. Stratech says that 12 ml of each preformulated reagent allows the user to complete more than 15 screens.

The method is suitable for hanging drop, sitting drop and sandwich drop crystallization methodologies. Ready-to-use reagents are formulated using high-purity salts, buffers and precipitants, with ultra-high-purity water and are sterile filtered.

■ Modes of synthesis

The new Expedite **nucleic acid synthesis system** from Millipore is designed to allow researchers to synthesize oligonucleotides in a few hours (*Reader Service No. 104*). With the company's 'alternating phase synthesis technology', two columns can run independently with fast cycle times of less than 4 min for one or two columns. In



DNA in half a day with the Expedite nucleic acid synthesis system.

addition, throughput can be increased when it is necessary to run more than one scale of synthesis. Millipore says that the microfluidic plate technology allows for reduced reagent consumption (less than 4.5 ml at the 0.05-, 0.2- and 1.0- μ mol scales). Additional features of the Expedite system include precalibrated injectors to eliminate manual flow rate; easy-to-use system software that offers a full sequence editor with mixed sites, storage for up to 64 sequences, optional trityl monitoring, and eight user profiles; and Expedite β -cyanoethyl chemistry which is designed to decrease deprotection time from 6 h to 15 min at 55 °C, or from 24 h to 2 h at room temperature.

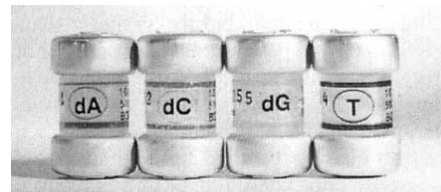
The new **DNA synthesis reagent kit** contains clearly labelled reagents in colour-coded bottles that correspond to labelling on the new Oligo 1000 DNA synthesizer from Beckman (*Reader Service No. 105*). The kit contains noncoupling reagents Oxidize (3 per cent I, in THF, pyridine, water), Deblock



Beckman's reagents — designed to simplify DNA synthesis.

(3 per cent TCA in DCM), Cap 1 (10 per cent Ac_2O in THF, lutidine) and Cap 2 (17 per cent NMI in THF). They are supplied in premeasured quantities for simultaneous depletion and replenishment when used on the Oligo 1000 DNA synthesizer. Non-coupling reagents are packaged under inert conditions to eliminate moisture and oxygen contamination for greater stability and long shelf life. Available separately, the coupling reagent Activate (0.5 M tetrazole/ CH_3CN) is used for converting the stable phosphoramidite into reactive species.

A complete line of **prepacked reaction columns for DNA and RNA synthesis** is now available from BioGenex Laboratories for use on synthesizers manufactured by Applied Biosystems (*Reader Service No. 106*). The columns are packed with DNA



DNA/RNA columns from BioGenex — just the job for ABI synthesizers.

and RNA monomers bound to CPG or WPS solid supports. Supplied with either 500 Å or 1,000 Å pore-size supports to accommodate oligomers of different lengths, the columns are also offered with either 1.0 μ mol or 0.2 μ mol of bound monomer to suit different scales of synthesis. In addition to DMT-deoxyribonucleoside-succinate-LCAA for DNA synthesis, columns are available with the corresponding base labile N-protection (N-PAC) for both RNA and DNA synthesis. The N-PAC protecting group provides both increased purity and higher yield of synthetic oligomer. The derivatization level

(loading) of the solid support, as determined by a DMT release assay, is reported on a quality assurance certificate that accompanies each shipment. ABI-compatible columns are precision machined from inert plastic resistant to the corrosive reagents and solvents used for automated synthesis.

■ Electrophoresis systems

According to Applied Biosystems, new protocols, chemistries, accessories and developments in the optics technology have enhanced the analytical capability and convenience of its Model 270A-HT capillary electrophoresis system (*Reader Service No. 107*). Protocols for isoelectric focusing and fraction collection and a new chemistry kit



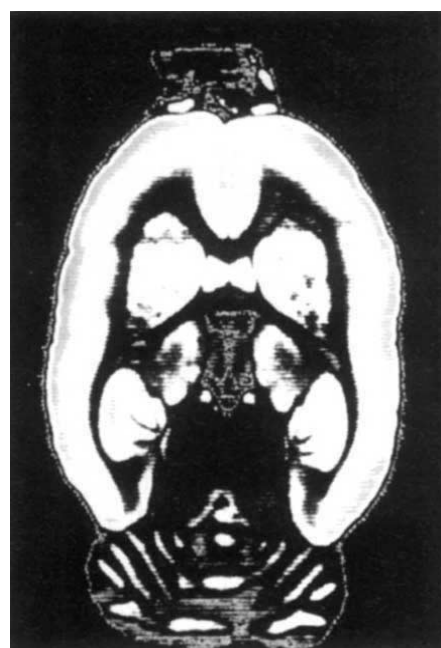
New protocols, chemistries, accessories and powerful optics technology have enhanced the analytical capability of the Model 270A-HT.

for molecular weight determination of proteins have expanded the utility of the system for research and methods development. In addition to new specialty applications kits, the Model 270A-HT is compatible with a variety of computer systems for data handling. ABI's Model 600 data system is a Macintosh-based program for data handling. So whether you are seeking added analytical information to complement HPLC analysis, or looking to maximize the separation power of electrophoresis (previously available only through slab-gel methods), ABI says that capillary electrophoresis offers convenience, high throughput, and precise quantitation using sample volumes as small as 5 µl.

The Model 491 Prep Cell from Bio-Rad purifies proteins, peptides and glycoproteins from complex mixtures by **continuous elution electrophoresis** (*Reader Service No. 108*). The system is designed to be used with conventional gel electrophoresis buffer systems and media. Using SDS-PAGE or native PAGE, the Model 491 Prep Cell can isolate specific components from complex mixtures containing micrograms to 200 mg of total sample. Bio-Rad says that up to 5 mg per band can be resolved. Using SDS-PAGE, the cell isolates molecules that differ in molecular weight by 2 per cent; using nondenaturing PAGE, it isolates molecules that differ in charge by 0.1 pH units. During a typical 3–7-h run, samples are electrophoresed through a cylindrical gel. Individual bands migrate off the bottom of the gel where they pass directly into the elution chamber for collection. As molecules are purified, they are collected in individual liquid fractions and are immediately available for sequence analysis, crystallization, the production of antibodies, studies on enzyme kinetics, or any other study where highly purified proteins are required.

■ Assorted reagents

A recently discovered **ligand that is selective for N-type calcium channels**, Conotoxin MVIIA, is now available in a radiolabelled form through Amersham (*Reader Service No. 109*). The new radiolabelled peptide is designed to complement the existing [¹²⁵I] Conotoxin GVIA, but with the additional benefit of enhanced reversible binding. [¹²⁵I] Conotoxin MVIIA shows a high affinity for binding in both rat and human neurocortical membranes, which allows competitive analysis of N-type calcium channel modulators. Amersham says that so specific is this radioligand that single sites can be labelled in a neuronal preparation. This specificity, in addition to its potential to be displaced by competitive modulators, makes Conotoxin MVIIA a useful tool in brain tissue receptor autoradiograph studies. In addition to Conotoxin MVIIA, Amersham is offering radiolabelled Galanin for use in radioimmunoassays and in receptor binding assays. Available in a freeze-dried formula-

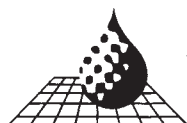


Conotoxin MVIIA — Amersham's new radiolabelled ligand for brain research.

tion, the new (3-[¹²⁵I] Iodotyrosyl)²⁶ Galanin has a radiochemical purity of greater than 90 per cent and a specific radioactivity of 74 Tbq mmol⁻¹ (2,000 Ci mmol⁻¹). Galanin, originally identified as a neuropeptide, is now known to be present in the pancreas and adrenal glands, as well as the peripheral and central nervous system. Galanin is also known to be a potent inhibitor of glucose-induced insulin release in several animal models, notably rat, mouse and dog.

Advanced Biotechnologies announces the availability of a new **monoclonal antibody to atrial natriuretic peptide (ANP)** — a 28-amino acid peptide produced by the cardiac atrium (*Reader Service No. 110*). The monoclonal, which has an IgG1 isotype, is presented as mouse ascites and has an affinity of 7 × 10¹⁰ l mol⁻¹. The company says that it has been used successfully in both radioimmunoassays and enzyme immunoassays. Atrial natriuretic peptide has potent diuretic effects. It is a vasodilator and

ADVERTISEMENT



Severn Biotech Ltd

Unit 23, Hoobrook Enterprise Centre,
Worcester Road, Kidderminster DY10 1HY

Custom DNA & RNA Oligo's

Probes, Primers, Antisense Oligo's, Ribonucleotides, 5' and 3' labelling with Biotin, Fluorescein, Rhodamine, Modified bases purification etc.

Complete Peptide Service

Small Scale Multiple Peptides 5mgs or larger scale 25mgs - 100mgs using L and D Amino Acids, screening grade or HPLC purified. Phosphorylated, Acetylated, Amidated and Protein Carrier Conjugated.

For a Fast Efficient Service Reply Today

Tel: (0562) 825286 Fax: (0562) 825284

CUSTOM SERVICES

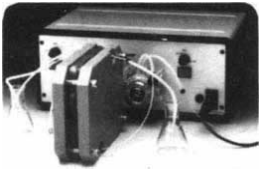
Comprehensive Custom Services for DNA, RNA and Peptides from SBL

- DNA Synthesis
- RNA Synthesis
- Multiple Peptide Synthesis (5mgs)
- Larger Scale Peptide Synthesis 25 - 100 mgs
- HPLC Purification
- Custom Conjugation to Carriers
- Peptide Sequencing

Reader Service No.21

ADVERTISEMENTS

LIPOSOMES
REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM
POB 126, D-8000 Munich 65, FRG

Reader Service No.23

Let the experts do the job

ABIMED now offers high quality custom synthesized peptides at very competitive prices. Call or FAX for further information.

- * Quantity 5-50 mg
- * Up to 25 residues
- * HPLC analysis included
- * Crude or purified
- * Mass spectrum available

ABIMED Analysen-Technik GmbH
PB 11 11, D-4018 Langenfeld, FRG. Phone: (+49)-2173-8905-0 FAX: (+49)-2173-8905-77

Reader Service No.6

**Highest Quality
Oligonucleotides**

Synthesised, purified and shipped in 3 days

£2.75 per base
(All inclusive)
0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit
King's College School of Medicine & Dentistry
Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

\$3.50/Base

**Custom DNA
Synthesis**

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

KEYSTONE
Laboratories, Inc.

Reader Service No.33

has inhibitory effects on the secretion of renin and aldosterone. Measurement of ANP may be a useful indicator in the monitoring and diagnosis of various cardiovascular diseases, including hypertension, cirrhosis and primary aldosteronism.

Multiple Peptide Systems is now offering a **new schedule for peptide synthesis** called 'Weekly Peptides' (Reader Service No. 111). The company says that every peptide ordered by 6.00 am Tuesday begins production by 6.00 am on Wednesday, every week, and that the semi-automated synthesis technology enables hundreds of peptides to be synthesized weekly. Included with every peptide synthesized at MPS is both HPLC and mass spectral analysis at no additional charge. Five purity levels are offered: unpurified, 55, 80, 95 and 98 per cent. All custom synthesized peptide sequences are confidential.

The BT7400 from Biotech Instruments is designed to be a low-cost **multiple peptide synthesizer** for use in applications where large numbers of peptides are required in small quantities, a typical yield being 5-10 µmol (Reader Service No. 112). The system consists of a polypropylene block containing 48 wells, into which a quantity of resin is added. All reagents and solvents are added to the wells using an Eppendorf pipette and cleavage is carried out on the unit, without having to transfer the resin to separate vials. The BT7400 is supplied with software that is designed to simplify both the synthesis strategy and the operation of the hardware. It allows the operator to input a protein sequence of up to 1,000 amino acids and store it to disk for future use. Peptides based on sequences of any length from the stored master sequence can be pulled down, or peptides of a given length can be selected. In this case, it is possible to overlap peptides by any number of amino acids. A fully automated version is also available.

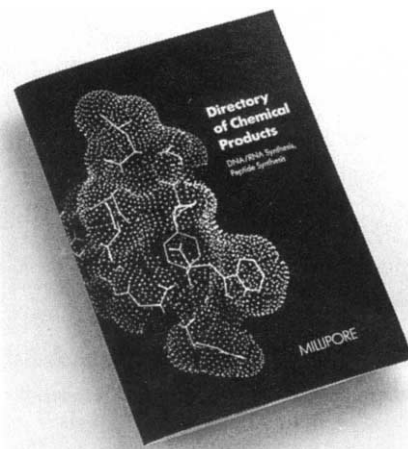
■ **Catalogues/applications guides**
Clontech introduces *OligoTechniques: Innovative Chemistry for Biological Applications*, a newsletter that sets out to describe novel **techniques and applications for nonisotopically labelling oligonucleotides** (Reader Service No. 113). The newsletter will be published three times a year and will cover topics such as antisense/triple helix and nonisotopic detection techniques.

Calbiochem-Novabiochem announces its new Novasyn **product line for peptide synthesis** (Reader Service No. 114). The line-up includes amino acid derivatives, resins and reagents in vials compatible with Applied Biosystems' Model 430A/431A and Millipore's 9050 synthesizers.

A comprehensive line of **systems and reagents for protein and DNA analysis** is represented in the new product catalogue

from Applied Biosystems (Reader Service No. 115). The 184-page volume features the latest technologies and products for genetic analysis, protein sequencing, peptide synthesis and bioseparations. In addition to details about the product, the catalogue features useful information about the selection of instruments and the application and use of reagents for DNA sequencing, polymerase chain reaction and nucleic acid synthesis.

The new Millipore **directory of chemical products** and price list is now available (Reader Service No. 116). The 52-page



Millipore's biosciences catalogue.

directory features the company's entire range of chemicals and reagents used in DNA/RNA synthesis and peptide synthesis. Other useful information like technical tables and molecular structures are also included. □

These notes have been compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in reader service number 100.

CORRECTION

The following product announcement, which appeared in a March issue of *Nature* (362, 189; 1993) should have read as follows: Cyberforce is Sera-Tech's new **liquid preparation developed for tissue culture purposes** that is made from native fetal bovine serum (FBS) and enhanced in its biological activities by defined components (Reader Service No. 117). Compared with native FBS, it can be used in drastically reduced concentrations of 0.1% to 1%. It renders serum-dependent cell culture systems (such as transformed cell lines, tumour cells, hybridomas and primary cells) to grow nearly serum free. No supplemental mitogens, hormones, cytokines, tumour-promoting components or vitamins have been added. Sera-Tech recommends a standard concentration of 0.3%. Cells with a higher need for FBS should be cultured with an additional 0.5% up to 2.0% FBS. The company says that with the low concentration of Cyberforce the total protein content of the medium is reduced to 0.04-0.4 mg ml⁻¹, and therefore the presence of undefined components is also reduced.